Unravelling the stellar Initial Mass Function of early-type galaxies with hierarchical Bayesian modelling

Dries, Matthijs

rijksuniversiteit
groningen

Unravelling the stellar Initial Mass Function of Early-Type Galaxies with hierarchical Bayesian modelling

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 6 april 2018 om 16:15 uur

door

Matthijs Dries

geboren op 10 februari 1986
te Emmen

Promotores
Prof. dr. S. C. Trager
Prof. dr. L. V. E. Koopmans

Beoordelingscommissie
Prof. dr. C. Conroy
Prof. dr. R. F. Peletier
Prof. dr. T. Treu

ISBN: 978-94-034-0465-3 (printed version)
ISBN: 978-94-034-0464-6 (electronic version)

Cover: Artist impression of the elliptical galaxy Messier 49 unravelling into a spectrum that can be used to infer the IMF of the galaxy
Design by: Caroline Boerema
Image credit: NASA and The Hubble Heritage Team (STScI/AURA) for the original image of M49 and the Digitized Sky Survey (DSS), STScI/AURA, Palomar/Caltech and UKSTU/AAO for the background star field

Acknowledgement: The work contained in this thesis was supported by an NWO grant (project number 614.001.208)

Printed by: Gildeprint - Enschede

Contents

Chapter 1

Introduction

On a dark moonless night in the countryside one can discern a band of light that stretches across the sky. Already since the ancient Greeks, this band of light has been known as the Milky Way. It was only in 1610 that Galileo Galilei for the first time pointed a telescope at the Milky Way and discovered that it actually consisted of many faint stars that are not resolved by the naked eye.

More than a century later, the French comet hunter Charles Messier continuously ran into a set of fuzzy objects that were actually not comets because they did not move across the sky. To prevent him and his colleagues from wasting time on these objects in their search for comets, he decided to make a list of these objects. This list is now known as the Messier Catalogue. In the next century, many more of these objects were discovered, by, amongst others, the German-British astronomer William Herschel. This led to the publication of the extensive New General Catalogue (NGC) by John Dreyer in 1888.

The origin of the so-called "spiral nebulae" found in these catalogues has long been the subject of debate. Of particular interest in this discussion was the question whether these objects originated in the Milky Way or had an extragalactic origin. A noticeable highlight in the discussion was the 1920 Great Debate between Harlow Shapley and Heber Curtis on the "scale of the Universe". It was, however, only in 1923 that the American astronomer Edwin Hubble finally resolved the debate. Using Cepheid variable stars, he managed to measure the distance to the Andromeda nebula and show that it is located far beyond the Milky Way. Nowadays, the Milky Way is known to be only one out of billions of other galaxies.

In the next decade Hubble continued his research into galaxies and developed a classification scheme for galaxies based on their visual appearance. This classification scheme is known as the Hubble sequence and divides galaxies into three main categories: ellipticals, lenticulars and spirals. Figure 1.1 gives a schematic overview of the Hubble sequence. Related to their location in the Hubble diagram, elliptical and lenticular galaxies are also referred to as early-type galaxies, whereas spirals are referred to as late-type galaxies.

Besides their morphological differences, early-type and late-type galaxies are also characterized by different colours. Late-type galaxies are in general bluish whereas early-type galaxies are yellowish-reddish. The different colours are related to the stars that constitute these galaxies. Spiral galaxies are actively forming stars and the most massive stars that are

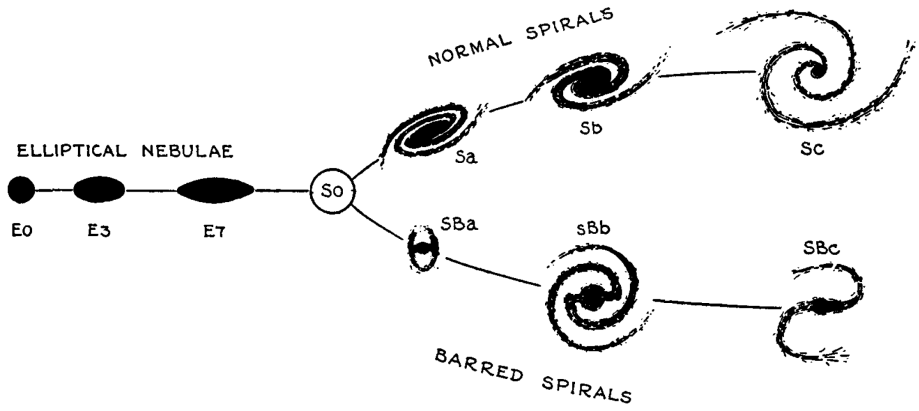

Figure 1.1 – Classification of galaxies in the Hubble tuning fork diagram. Image credit: Hubble (1936).

formed are hot and blue. These massive stars outshine their less massive, cooler and redder counterparts, making the galaxy appear blue. Elliptical galaxies, on the other hand, have very little ongoing star formation. The massive stars that make spiral galaxies blue have relatively short lifetimes, and hence these stars are no longer present in ellipticals. Therefore, the light of ellipticals is dominated by the long-lived redder population of lower-mass stars, giving ellipticals their characteristic yellow-red colour.

Low-mass stars are thought to be much more abundant than high-mass stars. Within the Milky Way, star counts of individual stars have given us a reasonable idea of how the number of stars changes as a function of stellar mass (Salpeter 1955; Kroupa 2001; Chabrier 2003). For more distant galaxies, stars are unresolved and these kind of measurements are not possible. The distribution of stellar masses in these galaxies is therefore often assumed to be the same as in the Milky Way. But is the stellar mass distribution really the same in all galaxies? And how can one reliably measure the distribution of stellar masses in unresolved galaxies? These questions are the central theme of this thesis, where we develop a Bayesian model for inferring the stellar mass distribution of unresolved galaxies. In the remaining of this chapter, I will give a general introduction into the formation and evolution of galaxies (Section 1.1), stellar population synthesis (Section 1.3), the initial mass function (IMF) (Section 1.4), and

Bayesian model comparison (Section 1.5). The outline of this thesis will be given in Section 1.6.

1.1 Formation and evolution of galaxies

Nowadays, the standard model of cosmological structure formation is provided by the Λ cold dark matter (ΛCDM) paradigm (Liddle 2003). The ΛCDM paradigm is a cosmological model that seeks to explain the expansion of the Universe and the formation and evolution of the structures within it. The ΛCDM paradigm assumes that the Universe started with a single event: the Big Bang.

Within the ΛCDM model, the energy content of the Universe is made of three different components: radiation, matter and dark energy. The matter component can be divided into a baryonic component (i.e. the ordinary matter of which stars and planets are made) and a (cold) dark matter component. Dark matter only interacts through gravity and the weak interaction, and although there are several indications of its existence, the exact nature of it is currently unknown.

During the first three minutes after the Big Bang, the Universe was very hot and the hydrogen, helium and lithium atoms that formed during primordial nucleosynthesis remained completely ionized. This implies that the Universe is effectively opaque to electromagnetic radiation, since all photons were almost instantaneously scattered by free electrons and radiation and baryonic matter were in thermal equilibrium. As the Universe expanded it cooled down, and roughly 380,000 years after the Big Bang, the temperature became low enough for protons and electrons to form neutral hydrogen (the most abundant element in the Universe). This is referred to as recombination. After recombination, the Universe became basically transparent to radiation, and baryonic matter decoupled from radiation. A relic of this surface of last scattering is the Cosmic Microwave Background (CMB) discovered by Penzias & Wilson (1965).

The very early Universe after the Big Bang is thought to consist of an almost isotropic and homogeneous density field. This initial density field contained very tiny quantum fluctuations, which according to the inflation model (Guth 1981) were amplified enormously during a rapid phase of exponential expansion. These fluctuations are thought to be the seeds of the variety of structures that we nowadays observe in the Universe and are

confirmed by small fluctuations ($\sim 10^{-5}$) in the CMB temperature (Jarosik et al. 2011; Planck Collaboration et al. 2016).

The energy content of the Universe is nowadays dominated by dark energy (68.6%) and dark matter (26.8%), whereas the contribution of baryonic matter is much smaller (4.9%) and in terms of energy density photons are almost negligible (Planck Collaboration et al. 2016).

Before recombination, baryonic matter is coupled to radiation and radiation pressure prevents structure growth in the baryonic density field. However, since dark matter does not interact with photons, the small initial fluctuations in the dark matter density field start to grow under the influence of gravity. While the cosmic density ρ_c is decreasing as a consequence of the expanding background, the density contrast $\delta = \rho/\rho_c - 1$ of overdense regions increases as it attracts matter from its surroundings.

Initially, structure growth is reasonably well described by linear perturbation theory (Peebles 1980). When the density contrast reaches a value of $\delta \approx 1$, it is no longer sufficient to describe the evolution of growing structures with linear perturbation theory and one has to use the full non-linear fluid equations. This phase is referred to as non-linear structure formation. Around the time that the density contrast reaches a value of $\delta \sim 1$, the overdensity reaches the point of turnaround, and it decouples from the expanding Universe and starts to collapse under the influence of its own gravity. The initial overdensity becomes a gravitationally bound object, which we refer to as a dark matter halo. When the dark matter halo has virialized, it has collapsed to approximately half its radius at turnaround.

After recombination, perturbations in the baryonic density field also start to grow under the influence of gravity. Since baryonic matter comprises only a relatively small fraction of the total mass in the Universe, it is expected to follow the dark matter distribution closely. The galaxies in our Universe are therefore thought to have formed in dark matter haloes.

According to the ΛCDM cosmology, structure formation is also hierarchical. In this scenario, the smallest-scale structures are the first to become non-linear and form dark matter haloes. As time progresses, these small dark matter haloes continuously merge to form more massive objects. The distribution of dark matter haloes in the Universe as a function of time may be derived from cosmological N-body simulations but there are also analytical formalisms such as the Press-Schechter formalism (Press & Schechter 1974).

As baryonic matter settles into hydrostatic equilibrium in dark matter haloes, it is able to lose energy through radiative cooling processes. At some point this may trigger Jeans instabilities (Jeans 1902) and lead to the formation of the first stars and galaxies. Since the infalling baryonic gas conserves its angular momentum, the first galaxies are expected to have a rotating disk shape.

Galaxies can grow through a series of subsequent merger events. Within that context, we distinguish between minor and major mergers. In the case of a minor merger, a small galaxy is "ingested" by a much larger galaxy, leaving the dynamical structure of the more massive progenitor largely intact. Massive disk galaxies such as the MW are thought to assemble their mass through a sequence of minor mergers. For a major merger, the galaxies that merge have comparable masses, and after a major merger the original dynamical structures of these galaxies are often destroyed and the newly formed object is dominated by random motions. Major mergers are thought to play an important role in the formation of elliptical galaxies. Another process that potentially plays an important role in the formation of galaxies are cold streams of gas that originate in the cosmic web and can provide fuel for star formation in a shock-heated dark matter halo (Birnboim & Dekel 2003; Dekel et al. 2009).

1.2 Elliptical galaxies

In this thesis we will focus on early-type galaxies (ETGs). The main reasons for this are that there are indications that ETGs have an IMF that is different from the MW IMF (see Section 1.4) and that the old stellar populations that characterize ETGs make it easier to infer their IMF.

Most of the galaxies in the Universe are spiral galaxies. Nevertheless, ETGs contain more than half of the stellar mass in the Universe (Bell et al. 2003; Gallazzi et al. 2008). ETGs are characterized by an ellipsoidal shape and a nearly smooth surface brightness distribution. For giant and midsized ellipticals the de Vaucouleurs' law (de Vaucouleurs 1948) provides a good description of the surface brightness profile. The de Vaucouleurs' profile is given by

$$\ln I(R) = \ln I_0 - kR^{1/4}, \tag{1.1}$$

in which R is the radius from the center and I_0 is the surface brightness at $R = 0$. More generally, the surface brightness profile of an ETG can be

described by a Sérsic profile (Sérsic 1963),

$$\ln I(R) = \ln I_0 - kR^{1/n}, \tag{1.2}$$

which for $n = 4$ reduces to the de Vaucouleurs' profile.

The main stellar component of an ETG is a population of old and red stars. Historically, ETGs are thought to contain little gas and dust and almost no ongoing star formation. However, there is evidence that some ETGs do contain gas and show signs of a small amount of residual star formation (Young et al. 2011; Crocker et al. 2011). In the densest environments of the Universe, elliptical galaxies are more common than in less dense environments. This is known as the morphology-density relation (Dressler 1980), and it provides important clues on the formation and evolution of galaxies.

There are several empirical scaling relations between the properties of elliptical galaxies. One of the most important scaling relations is the fundamental plane (Djorgovski & Davis 1987; Dressler et al. 1987), which relates the effective radius $R_{\rm eff}$ (i.e. the radius at which half of the total light of the system is emitted), the central velocity dispersion σ_c and the average surface brightness within the effective radius $\langle I_{\rm eff}\rangle$ of the galaxy as

$$R_{\rm eff} \propto \sigma_c^a \times \langle I_{\rm eff}\rangle^b. \tag{1.3}$$

By measuring the velocity dispersion σ_c and the average surface brightness within the effective radius $\langle I_{\rm eff}\rangle$, the fundamental plane can be used to determine the effective radius $R_{\rm eff}$. The physical size of the effective radius is not directly observable, but by combining it with the angular size of the galaxy, the fundamental plane allows one to determine distances to ETGs.

1.3 Stellar population synthesis

In the study of unresolved galaxies, spectra or spectral energy distributions of these galaxies are an essential source of information. Many of the physical processes that take place in these galaxies leave an imprint on the spectra that we observe. Modelling these spectra with a stellar population synthesis (SPS) model allows us to study the physical processes in these galaxies and to infer properties of these galaxies that are not directly observable. In this section I will give a basic overview of SPS models. For a more complete

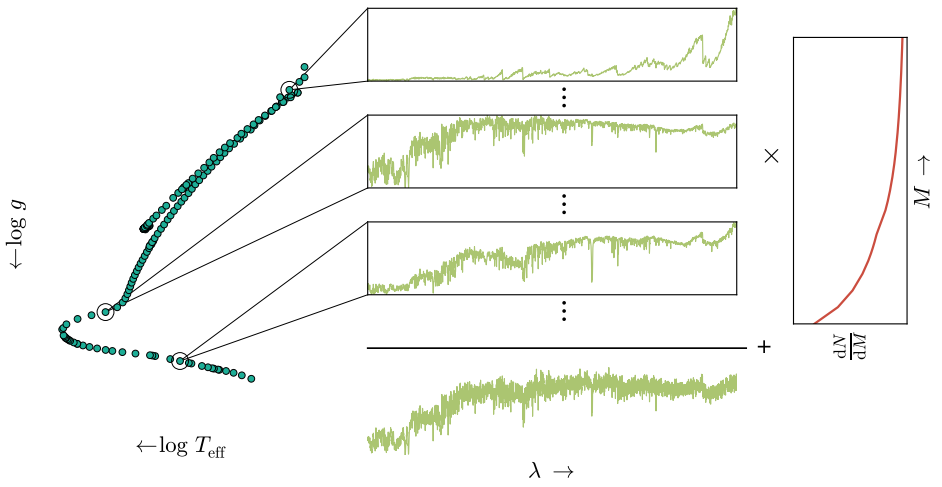

Figure 1.2 – Schematic overview of the construction of a model spectrum for an SSP.

overview I refer the reader to the excellent reviews by Walcher et al. (2011) and Conroy (2013).

1.3.1 Single Stellar Populations

A single stellar population (SSP) is a population of stars with the same age and the same chemical composition. Creating a model spectrum of an SSP required three basic ingredients: stellar evolution (isochrones), a stellar library and an IMF. An isochrone of a given age and metallicity provides the stellar parameters of the stars that are present in an SSP. When we combine these stellar parameters with a stellar library, this allows us to create a spectrum for each of the stars in the isochrone. Finally, the IMF prescribes the number of stars that we have for each of the stars in the isochrone. Multiplying each of the isochrone spectra with the number of stars that are present in an SSP according to the IMF allows us to create a model spectrum for the SSP. This process is schematically shown in Figure 1.2. Mathematically, the construction of an SSP model spectrum can be summarized as

$$f_{\rm SSP} = \int_{m_{\rm low}}^{m_{\rm up}(t)} f_{\rm star}[T_{\rm eff}(M), \log g(M)|t, Z]\xi(M)dM, \tag{1.4}$$

in which $f_{\rm star}[T_{\rm eff}(M), \log g(M)|t, Z]$ is the spectrum of a star with effective temperature $T_{\rm eff}$ and surface gravitity $\log\ g(M)$ for a given age t and metallicity Z. The IMF in the above relation is represented by $\xi(M) \equiv {\rm d}N/{\rm d}M$ and the integral is evaluated from the lowest mass stars $m_{\rm low}$ to the (time-dependent) highest mass stars $m_{\rm up}(t)$ (this upper limit is set by our knowledge of stellar evolution). Usually the lower limit is set to the hydrogen burning limit ($m_{\rm low} \approx 0.1$ M$_\odot$) and the initial upper limit is taken to be $m_{\rm up} \approx 100$ M$_\odot$.

The relation between the surface gravity, mass and effective temperature for a stellar population of a certain age and metallicity is defined by isochrones. The individual stellar spectra $f_{\rm star}[T_{\rm eff}(M), \log\ g(M)|t, Z]$ are taken from either theoretical models or empirical stellar libraries. We now discuss the isochrones and stellar libraries in more detail. The IMF is discussed in more detail in Section 1.4.

With numerical models of stellar evolution it is possible to calculate the evolutionary track of a star with a given mass and metallicity Z. An evolutionary track describes how a star evolves in the Hertzsprung-Russel (HR) diagram as a function of age. By calculating evolutionary tracks for a wide range of initial masses, e.g. $M = 0.1 - 100$ M$_\odot$, one can construct a library of evolutionary tracks. Evolving each of these tracks for a given metallicity Z to the same age t and combining the parameters of these tracks gives us an isochrone of age t and metallicity Z. One such isochrone defines the stellar parameters of the stars that are present in an SSP.

Nowadays, different sets of homogeneous evolutionary tracks and isochrones are available. Examples are the Parsec isochrones (Bertelli et al. 1994; Girardi et al. 2000; Marigo et al. 2008), the Geneva isochrones (Schaller et al. 1992; Meynet & Maeder 2000), the Dartmouth isochrones (Dotter et al. 2008), the BaSTI isochrones (Pietrinferni et al. 2004; Cordier et al. 2007) and the Lyon isochrones (Chabrier & Baraffe 1997; Baraffe et al. 1998). These isochrones may have different assumptions with respect to the input physics that was used to calculate the evolutionary tracks, and there are several uncertainties that may eventually affect the SSP model spectra that are created with these isochrones. Among these uncertainties are the use of one-dimensional stellar evolution codes, the adopted level of convective overshooting, the modelling of stellar rotation, the interaction of close binary stars, mass loss, and the modelling of the thermally pulsating asymptotic giant branch (TPAGB). Since there is not a single isochrone model that covers all stellar masses, metallicities and evolutionary stages,

it is quite common to combine different sets of isochrones. However, this is not straightforward due to the different assumptions made in different models.

To subsequently convert the stellar parameters of an isochrone into a set of stellar spectra requires the use of a stellar library. Stellar libraries can be either theoretical or empirical. Theoretical libraries (e.g. Kurucz 1992; Coelho et al. 2005; Husser et al. 2013) have the advantage of providing a dense coverage of parameter space and having a high spectral resolution. Important uncertainties in theoretical spectra are the treatment of convection, incomplete lists of atomic and molecular lines, uncertain strengths and central wavelengths of molecular and atomic lines, and insufficient knowledge of the molecular partition function. Since empirical libraries are based on real stars, they by definition suffer less from these uncertainties. The main uncertainties of empirical libraries are determined by observational constraints. Moreover, empirical libraries have the disadvantage of not completely covering the parameter space of the isochrones (especially for the late evolutionary stages and for lower metallicities). Another source of uncertainty for empirical libraries is the determination of the stellar parameters of the stars in the library. Examples of empirical libraries are STELIB (Le Borgne et al. 2003), ELODIE (Prugniel & Soubiran 2001, 2004; Prugniel et al. 2007), INDO-US (Valdes et al. 2004), MILES (Sánchez-Blázquez et al. 2006), IRTF (Rayner et al. 2009) and XSL (Chen et al. 2011).

1.3.2 Composite Stellar Populations

Modelling the spectrum of a real galaxy is further complicated by the fact that a galaxy is not an SSP but has an extended star formation history (SFH). Therefore, a galaxy is a composite stellar population (CSP). Moreover, the spectrum of a galaxy that we observe is affected by dust.

There are various indicators that may be used to probe the star formation rate of a galaxy. Examples of these indicators are the total UV luminosity, infra-red indicators based on the processing of starlight by dust and emission lines from ionized gas (Kennicutt & Evans 2012). However, to create an accurate model for the spectrum of a CSP requires an integration over the SFH of the CSP. Classically, the SFH of a galaxy has been assumed to be exponentially declining, resulting in so-called τ-models for which SFR $\propto e^{-t/\tau}$ (Papovich et al. 2001; Shapley et al. 2005). Other

parameterizations of the SFH include e.g. inverted τ-models (Maraston et al. 2010; Pforr et al. 2012) and delayed τ-models for which SFR $\propto te^{-t/\tau}$ (Sandage 1986; Lee et al. 2010). Besides these simple parameterizations it is also possible to use a library of SFHs from hydrodynamical simulations or semi-analytical models of galaxy formation (Finlator et al. 2007; Pacifici et al. 2012) or to use nonparametric SFHs (Ocvirk et al. 2006).

The observed spectrum of a galaxy can be affected by dust in a number of ways. First of all, dust along the line of sight to the observed galaxy causes extinction. An observed spectrum may be corrected for this by applying an extinction correction (Cardelli et al. 1989; Calzetti 2001). Secondly, in galaxies that contain dust, light at shorter wavelengths may be absorbed by the dust and re-emitted at longer wavelengths (in particular the infra-red). For ETGs this is in general not considered to be an important process because the amount of dust in ETGs is relatively low. Finally, asymptotic giant branch (AGB) stars are characterized by mass-loss and this mass-loss can potentially be dust-rich (Bedijn 1987) and therefore affect the spectra of AGB stars. It is not straightforward to model this process and most SPS models do not take it into account. Nevertheless, it might be important because AGB stars can contribute significantly to the total luminosity of a galaxy (Kelson & Holden 2010; Villaume et al. 2015).

Taking into account the SFH, the model spectrum $f_{\rm CSP}$ of a CSP may be written as

$$f_{\rm CSP}(t) = \int_{t'=0}^{t'=t} \int_{Z=0}^{Z=Z_{\rm max}} {\rm SFR}(t-t')\, P(Z, t-t')\, f_{\rm SSP}(t', Z)\, dt'\, dZ. \quad (1.5)$$

In this equation, SFR$(t-t')$ is the SFR at time $t-t'$ representing the SFH, $P(Z, t-t')$ is the time-dependent metallicity distribution function, and $f_{\rm SSP}$ is the spectrum of an SSP as calculated in Equation 1.4. Note that this equation neglects the effects of dust attenuation on the observed galaxy spectrum.

1.3.3 Population Synthesis models

Among the first population synthesis models were the models developed by Tinsley (1968), Spinrad & Taylor (1971) and Faber (1972). The models of Spinrad & Taylor (1971) and Faber (1972) tried to determine the individual contributions of stars in the HR diagram to integrated spectra but suffered from a lack of astrophysical constraints. Tinsley (1968) introduced the idea

of evolutionary populations synthesis in which theories of stellar evolution are used to put constraints on the available range of stellar templates in the HR diagram.

Nowadays, a variety of different population synthesis models have been developed (e.g. Bruzual & Charlot 2003; Le Borgne et al. 2003; Maraston 2005; Conroy & van Dokkum 2012a; Vazdekis et al. 2015) to infer the properties of unresolved galaxies. Every model has its own set of ingredients and parameters. Examples of such variety includes the isochrones that are used, the stellar library that is used, the parameterization of the IMF and modelling the spectrum as an SSP or not. The SFH of ETGs is expected to be different than that of the Milky Way, and as a consequence the chemical composition of ETGs is expected to be non-solar. Some population synthesis models take this effect into account by using response functions for variable abundance patterns. Another important aspect of a population synthesis model is how the models are fitted to the data. In what we described so far, the main focus has been the construction of a model spectrum for an SSP or CSP. Yet, ultimately we would like to use these models to extract information from observed galaxy spectra which requires some sort of a fitting routine.

1.4 The initial mass function

The IMF is a probability distribution that describes the distribution of stellar masses of the stars that form in a star formation event. The first attempt to measure the IMF was made by Salpeter (1955). Although often referred to as measuring the IMF, most of the time it is actually the present-day mass function (PDMF) that is measured. The PDMF includes the effects of stellar evolution and the SFH.

Salpeter (1955) derived the IMF for stars with masses between 0.4 and 10 $M_\odot$ in the solar neighbourhood and determined that it could be described with a single power law as

$$\xi(M) \equiv \frac{\mathrm{d}N}{\mathrm{d}M} \propto M^{-\alpha}, \tag{1.6}$$

with $\alpha = 2.35$. Later on it was shown that for the MW below 0.5 $M_\odot$ the IMF becomes flatter and that a broken power law (Kroupa et al. 1993; Kroupa 2001) or a combination of a power law with a lognormal distribution (Chabrier 2003) provides a more accurate description of the MW IMF. In

fact, these two distributions can be very similar and hard to distinguish, as shown by Dabringhausen et al. (2008).

Measurements of the IMF in the MW are mostly based on star counts. These measurements are complicated by the different volumes that are probed for different stellar masses, by the degeneracy with the SFH of the MW (Elmegreen & Scalo 2006) and by the effect of binaries on the inferred stellar mass spectrum (Metchev & Hillenbrand 2009; Goodwin & Kouwenhoven 2009). Studies of the IMF in different environments of the MW seem to indicate that the Galactic IMF is universal (Bastian et al. 2010).

For galaxies beyond the Local Group, direct star counts are not possible any more because the light of these galaxies is spatially unresolved. To determine the IMF of these systems one has to use different techniques. One of these techniques is the analysis of the spectrum of such a galaxy with a population synthesis model. Inferring the IMF and in particular the low-mass end of the IMF from the spectrum of a galaxy is not straightforward, however. Dwarfs with masses below 0.5 $M_{\odot}$ contribute only a few percent to the integrated light of an old SSP. Nevertheless, these dwarfs contribute significantly to the total stellar mass of the SSP. In young stellar populations the light is dominated by the young stars that are still present in those populations and therefore the relative contribution of low-mass dwarfs to the integrated spectrum is even smaller. The situation is complicated further by the fact that the spectra of low-mass stars and K- and M-giants are very similar. However, there are a number of (gravity-sensitive) features in the spectrum that allow us to distinguish between dwarfs and giants (Wing & Ford 1969; Faber & French 1980; Gorgas et al. 1993; Worthey et al. 1994; Schiavon et al. 1997a,b; Cenarro et al. 2003; Schiavon 2007; Spiniello et al. 2012).

There is increasing evidence from spectroscopic studies of ETGs that the IMF is not universal. These studies suggest that ETGs with a higher mass/velocity dispersion/metallicity contain a higher fraction of low-mass stars (Conroy & van Dokkum 2012b; Spiniello et al. 2012; Ferreras et al. 2013; La Barbera et al. 2013; Martín-Navarro et al. 2015b). Within these galaxies, there appears to be a gradient that suggests that the IMF in the centre of the galaxy is more bottom-heavy than it is for more outer radii (Martín-Navarro et al. 2015a; van Dokkum et al. 2016). Combined studies of population synthesis, dynamics and/or gravitational lensing also find higher mass-to-light ratios for these galaxies than the MW-value (Treu et al.

2010; Graves & Faber 2010; Cappellari et al. 2012; Conroy & van Dokkum 2012b; Lyubenova et al. 2016). In contrast to these findings, Smith et al. (2015) and Newman et al. (2016) have found a number of nearby ETGs that are consistent with the MW IMF but the spread in slopes is still large.

In many astrophysical studies the IMF is a fundamental input quantity that is often assumed to be universal and similar to the MW IMF. For example, mass-to-light ratios of galaxies, the chemical evolution of galaxies, the energy balance of the interstellar medium, the population of stellar remnants, and the supernovae rate are all related to the IMF. If the IMF is not universal, this may affect any of the galaxy properties that are derived under the assumption of a universal IMF (Ferré-Mateu et al. 2013; Fontanot et al. 2017). Accurately measuring the shape of the IMF and possible variations of the IMF shape with other galaxy properties is therefore of crucial importance.

If one finds a correlation between for example the velocity dispersion of a galaxy and the slope of the IMF, this does not imply that there is a causal relation between these properties. In fact, it is highly unlikely that variations of the IMF are driven by global properties of the galaxy. Instead, one would expect that variations of the IMF are related to differences in the local properties of the clouds in which stars form, such as the pressure or the Mach number (Hennebelle & Chabrier 2008; Krumholz 2011; Hopkins 2012, 2013). These local properties may in turn reflect different typical circumstances of star formation in different galaxies which would then explain the empirical relation between for example IMF slope and velocity dispersion. Star formation is, however, a complex physical process that involves an interplay between gravity, turbulence, radiation, magnetic fields and the chemical composition of the gas. It is challenging to provide a theoretical framework work for this and the origin of the IMF in the process of star formation is not yet completely understood (Krumholz 2014; Offner et al. 2014). In that context, an accurate determination of how the IMF varies with other galaxy properties provides important constraints for any theory of star formation.

1.5 Bayesian model comparison

In the previous sections we have described how to construct a model for the spectrum of a stellar population, and the role of the IMF in the construction of such a model. Yet the ultimate goal is to extract the IMF from an

observed spectrum and to select the model that best fits the data. A statistically sound way to do this is through Bayesian model comparison.

Bayesian probability theory is based on two rules (Cox 1946) that may be used for manipulating conditional probabilities:

$$p(H|I) + p(\overline{H}|I) = 1 \tag{1.7}$$

and

$$p(H, D|I) = p(D|H, I)p(H|I) = p(H|D, I)p(D|I) \tag{1.8}$$

where H denotes the hypothesis, D denotes the data, $\overline{H}$ stands for not H, I represents any relevant background information and $|$ should be read as "given that". The first rule states that we are dealing with exclusive probabilities: adding the probability of H and the probability of not H is equal to one, where one means "true". The second rule is the product rule and may be used to derive Bayes' theorem:

$$p(H|D, I) = \frac{p(D|H, I)p(H|I)}{p(D|I)}, \tag{1.9}$$

where $p(H|D, I)$ represents the posterior probability, $p(D|H, I)$ the likelihood function, $p(H|I)$ the prior and $p(D|I)$ is usually referred to as the evidence. The strength of Bayes' theorem lies in its ability of transforming the likelihood, which is usually easier to calculate, into a posterior probability, which is the term we are actually interested in.

An important concept in Bayesian analysis is that of marginalization. The marginalization equation is given by

$$p(x|I) = \int_{-\infty}^{+\infty} p(x, y|I)dy = \int_{-\infty}^{+\infty} p(x|y, I)p(y|I)dy. \tag{1.10}$$

Marginalization may be used to include uncertainties from a nuisance model parameter y by integrating the conditional probability $p(x|y, I)$ multiplied by the probability distribution of the nuisance parameter y over the complete parameter space of y to obtain the probability distribution of the parameter we are interested in (x in this case).

In the process of data modelling, we can distinguish at least two levels of inference. The first level of inference assumes that a given model M_0 is true and then fits the free parameters of that model to the data. We refer to the

first level of inference as parameter estimation. When Bayesian methods are used for parameter estimation, the results are in general very similar to those derived with classical statistical methods. The real difference between classical statistical methods and Bayesian methods is found in the second level of inference, where different models are compared in the light of the probability of the data (i.e. the evidence).

Models that are more complex can nearly always fit the data better. Therefore, comparing different models on the basis of the maximum likelihood one will nearly always select the more complex model. However, this more complex model is potentially over-parameterized. For example, consider the data in Figure 1.3 that we fit with two different models. The first model M_0 is a straight line with two free parameters, the second model M_1 a tenth-order polynomial with eleven free parameters. Although the second model provides a better fit to the data, the chances are that you think the first model is more appealing. Bayesian model comparison allows us to quantify the comparison between different models in the light of the data. Hence, Bayesian model comparison automatically embodies Occam's razor, stating that we should select the "simplest" model that fits the data.

Suppose that we want to compare two models M_0 and M_1 to describe a given set of data D. The free parameters in the model are θ_0 and θ_1. At the first level of inference, we infer the posterior probability of the model parameters θ_i. According to Bayes' theorem, this probability distribution can be written as

$$p(\theta_i|D, M_i) = \frac{p(D|\theta_i, M_i)p(\theta_i|M_i)}{p(D|M_i)}. \tag{1.11}$$

For the first level of inference the normalizing constant, or the evidence, is not important but for the second level of inference this term is very important.

When we compare the models M_0 and M_1, we once again apply Bayes' theorem to calculate the odds ratio

$$\frac{p(M_0|D)}{p(M_1|D)} = \frac{p(D|M_0)p(M_0)}{p(D|M_1)p(M_1)}, \tag{1.12}$$

where the normalizing constants cancel out in the numerator and denominator. Assuming that we assign equal prior probabilities to the two models, we can compare the models on the basis of the ratio $p(D|M_0)/p(D|M_1)$ which are the normalizing constants of Equation 1.11. In other words, to

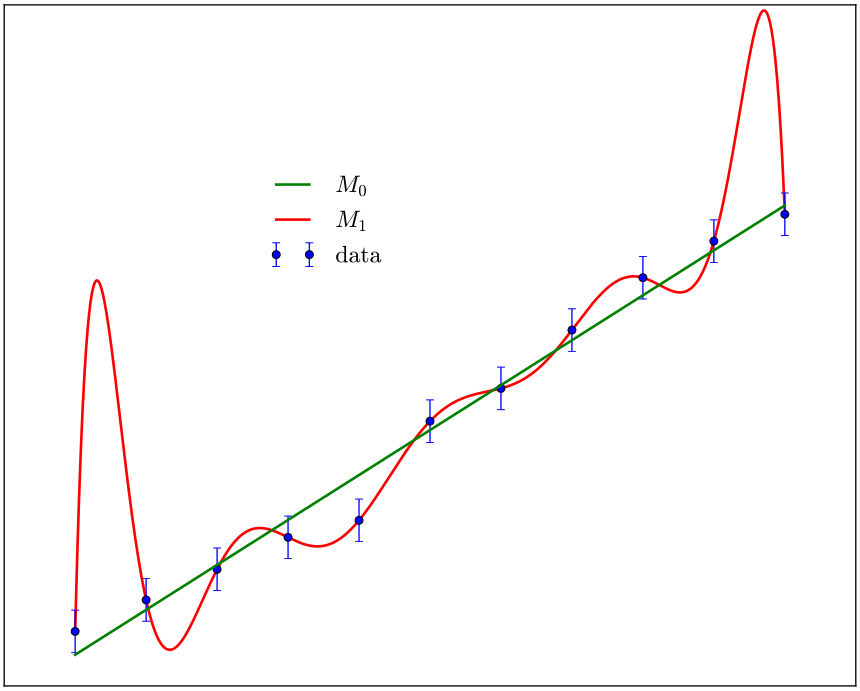

Figure 1.3 – More complex models always fit the data better. According to Occam's razor one should select the simplest model that fits the data.

compare models M_0 and M_1 we need to evaluate the evidence $p(D|M_0)$ for model M_0 and compare it to the evidence $p(D|M_1)$ for model M_1.

To understand how Bayesian model comparison automatically embodies Occam's razor it is insightful to look at the simple case of a model M with one free parameter θ (see also MacKay 1992, for a more extensive discussion). The evidence for that model may be written as

$$p(D|M) = \int p(D, \theta|M)\mathrm{d}\theta = \int p(D|\theta, M)p(\theta|M)\mathrm{d}\theta. \tag{1.13}$$

The first term in the integral is the likelihood of the data given the model and its free parameters, i.e., how well does the model fit the data. The second term is the prior for the probability distribution function of θ. If

we assume that θ lies between $\theta_{\min}$ and $\theta_{\max}$ with a uniform prior, we may write

$$p(\theta|M) = \frac{1}{\theta_{\max} - \theta_{\min}} \quad \text{for } \theta_{\min} \leq \theta \leq \theta_{\max}, \tag{1.14}$$

and we have

$$p(D|M) = \frac{1}{\theta_{\max} - \theta_{\min}} \int_{\theta_{\min}}^{\theta_{\max}} p(D|\theta, M) \mathrm{d}\theta. \tag{1.15}$$

Assuming that $p(D|\theta, M)$ can be described by a Gaussian centred at θ_0 with standard deviation σ_θ that lies well within the prior range of θ, the integral evaluates to $\sqrt{2\pi}\sigma_\theta$ so that

$$p(D|M) = p(D|\theta_0, M) \times \frac{\sqrt{2\pi}\sigma_\theta}{\theta_{\max} - \theta_{\min}} = L_{\max} W. \tag{1.16}$$

The second factor in this equation is referred to as the Occam factor and penalizes the model M for having parameter θ. Equation 1.16 shows that the evidence can be written as the product of the maximum likelihood $L_{\max}$ times the Occam factor W. For models with one parameter, the Occam factor is the ratio of the width of the likelihood over the width of the prior. When the model includes multiple parameters this may be generalized to the ratio of the prior volume that provides a good fit to the data by the total volume that is accessible by the prior. For models with a broader prior or more parameters, the accessible prior volume is larger and hence the Occam factor becomes smaller and more complex models are automatically penalized. The evidence is therefore always a trade-off between the quality of the fit to the data and the model complexity. As long as an extra parameter improves the fit to the data sufficiently, the decrease of the Occam factor is compensated by an increase of the likelihood. However, if the likelihood only increases by a small amount when an additional parameter is included, Bayesian model comparison will prefer the simpler model with fewer parameters, because it has a higher probability to produce the data.

1.6 This thesis

The IMF is a fundamental quantity in the study of galaxy formation and evolution. A non-universal IMF has important consequences for many astrophysical studies and provides constraints for theories of star formation. Although there is increasing evidence that the IMF is not universal, the exact shape of the IMF and the scale of these IMF variations remain uncertain.

Inferring the IMF from the spectrum of a galaxy with a population synthesis model is not straightforward. Nowadays there are many different SPS models with a variety of different ingredients to choose from. Possible degeneracies between these ingredients and the inferred IMF as well as selecting which model one should use requires a solid statistical framework.

In this thesis we develop a hierarchical Bayesian framework for population synthesis that is specifically designed for inferring the IMF of unresolved stellar populations. However, the framework that we develop is more general and can be used to infer other properties of unresolved stellar populations as well. Within the model that we develop we use a parameterized IMF prior to regulate a more direct inference of the IMF shape. This direct inference gives more freedom to the model and allows the model to deviate from parameterized models when demanded by the data. The Bayesian framework of the model that we develop makes it well-suited for model comparison and allows us to objectively compare different ingredients of SPS models.

In Chapter 2 we discuss the general setup of the hierarchical Bayesian framework for SSPs. We combine the model with stellar templates that are based on the MILES library and use these templates to create a set of mock SSP spectra. Then we apply the model to these mock spectra and show that we can reconstruct the input parameters of the mock SSPs. However, when we apply the model to mock SSPs created with a different population synthesis code, this may introduce a bias on the inferred IMF parameters.

Many spectroscopic IMF studies of ETGs assume that an ETG can be modelled as an SSP. To test this assumption, in Chapter 3 we extend the SSP model of Chapter 2 to CSPs. We assume that we can model a CSP as a combination of SSPs and use the Bayesian evidence to discriminate between models with different numbers of SSPs. Using SFHs based on semi-analytic models, we create a number of realistic CSP mock spectra with an IMF that varies as a function of the velocity dispersion of the galaxy. Then we try

to reconstruct the IMF of these mock CSPs by using a variable number of SSPs in the fit.

One other motivation for starting the research described in this thesis has been the development of the X-shooter Spectral Library (XSL). The stars in XSL have been selected to provide a good coverage of the HR diagram and the spectra in the library extend all the way from the UV to the NIR. This makes XSL an ideal library for population synthesis studies of the IMF.

In Chapter 4, we combine the VIS arm of XSL with the MILES library to create the MIX stellar population models. Moreover, we extend the model developed in Chapter 2 and 3 to include various response functions to account for variations in the abundance pattern. Then, we combine the hierarchical Bayesian framework that we developed with the MIX stellar population models and apply this to a set of stacked *SDSS* spectra binned by velocity dispersion. When we fit these *SDSS* spectra, we include multiple SSPs, response functions of various elements, two different isochrone sets, two different regularization schemes and two different parameterizations of the IMF prior, and we compare different model ingredients on the basis of the Bayesian evidence.

References

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403
Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339
Bedijn, P. J. 1987, A&A, 186, 136
Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289
Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., & Nasi, E. 1994, A&AS, 106, 275
Birnboim, Y. & Dekel, A. 2003, MNRAS, 345, 349
Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D. 2001, PASP, 113, 1449
Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2012, Nature, 484, 485
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Cenarro, A. J., Gorgas, J., Vazdekis, A., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 339, L12
Chabrier, G. 2003, PASP, 115, 763
Chabrier, G. & Baraffe, I. 1997, A&A, 327, 1039
Chen, Y., Trager, S., Peletier, R., & Lançon, A. 2011, Journal of Physics Conference Series, 328, 012023
Coelho, P., Barbuy, B., Meléndez, J., Schiavon, R. P., & Castilho, B. V. 2005, A&A, 443, 735
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C. & van Dokkum, P. 2012a, ApJ, 747, 69
Conroy, C. & van Dokkum, P. G. 2012b, ApJ, 760, 71
Cordier, D., Pietrinferni, A., Cassisi, S., & Salaris, M. 2007, AJ, 133, 468
Cox, R. 1946, Am. J. Phys., 14, 1
Crocker, A. F., Bureau, M., Young, L. M., & Combes, F. 2011, MNRAS, 410, 1197
Dabringhausen, J., Hilker, M., & Kroupa, P. 2008, MNRAS, 386, 864
de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247
Dekel, A., Birnboim, Y., Engel, G., et al. 2009, Nature, 457, 451
Djorgovski, S. & Davis, M. 1987, ApJ, 313, 59
Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89
Dressler, A. 1980, ApJ, 236, 351
Dressler, A., Lynden-Bell, D., Burstein, D., et al. 1987, ApJ, 313, 42

Elmegreen, B. G. & Scalo, J. 2006, ApJ, 636, 149
Faber, S. M. 1972, A&A, 20, 361
Faber, S. M. & French, H. B. 1980, ApJ, 235, 405
Ferré-Mateu, A., Vazdekis, A., & de la Rosa, I. G. 2013, MNRAS, 431, 440
Ferreras, I., La Barbera, F., de la Rosa, I. G., et al. 2013, MNRAS, 429, L15
Finlator, K., Davé, R., & Oppenheimer, B. D. 2007, MNRAS, 376, 1861
Fontanot, F., De Lucia, G., Hirschmann, M., et al. 2017, MNRAS, 464, 3812
Gallazzi, A., Brinchmann, J., Charlot, S., & White, S. D. M. 2008, MNRAS, 383, 1439
Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, A&AS, 141, 371
Goodwin, S. P. & Kouwenhoven, M. B. N. 2009, MNRAS, 397, L36
Gorgas, J., Faber, S. M., Burstein, D., et al. 1993, ApJS, 86, 153
Graves, G. J. & Faber, S. M. 2010, ApJ, 717, 803
Guth, A. H. 1981, Phys. Rev. D, 23, 347
Hennebelle, P. & Chabrier, G. 2008, ApJ, 684, 395
Hopkins, P. F. 2012, MNRAS, 423, 2037
Hopkins, P. F. 2013, MNRAS, 433, 170
Hubble, E. P. 1936, Realm of the Nebulae
Husser, T.-O., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6
Jarosik, N., Bennett, C. L., Dunkley, J., et al. 2011, ApJS, 192, 14
Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1
Kelson, D. D. & Holden, B. P. 2010, ApJL, 713, L28
Kennicutt, R. C. & Evans, N. J. 2012, ARA&A, 50, 531
Kroupa, P. 2001, MNRAS, 322, 231
Kroupa, P., Tout, C. A., & Gilmore, G. 1993, MNRAS, 262, 545
Krumholz, M. R. 2011, ApJ, 743, 110
Krumholz, M. R. 2014, Phys. Rep., 539, 49
Kurucz, R. L. 1992, in IAU Symposium, Vol. 149, The Stellar Populations of Galaxies, ed. B. Barbuy & A. Renzini, 225
La Barbera, F., Ferreras, I., Vazdekis, A., et al. 2013, MNRAS, 433, 3017
Le Borgne, J.-F., Bruzual, G., Pelló, R., et al. 2003, A&A, 402, 433
Lee, S.-K., Ferguson, H. C., Somerville, R. S., Wiklind, T., & Giavalisco, M. 2010, ApJ, 725, 1644
Liddle, A. 2003, An Introduction to Modern Cosmology (Wiley)

Lyubenova, M., Martín-Navarro, I., van de Ven, G., et al. 2016, MNRAS, 463, 3220
MacKay, D. J. C. 1992, Neural Computation, 4, 415
Maraston, C. 2005, MNRAS, 362, 799
Maraston, C., Pforr, J., Renzini, A., et al. 2010, MNRAS, 407, 830
Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883
Martín-Navarro, I., Barbera, F. L., Vazdekis, A., Falcón-Barroso, J., & Ferreras, I. 2015a, MNRAS, 447, 1033
Martín-Navarro, I., Vazdekis, A., La Barbera, F., et al. 2015b, ApJL, 806, L31
Metchev, S. A. & Hillenbrand, L. A. 2009, ApJS, 181, 62
Meynet, G. & Maeder, A. 2000, A&A, 361, 101
Newman, A. B., Smith, R. J., Conroy, C., Villaume, A., & van Dokkum, P. 2016, ArXiv e-prints
Ocvirk, P., Pichon, C., Lançon, A., & Thiébaut, E. 2006, MNRAS, 365, 46
Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2014, Protostars and Planets VI, 53
Pacifici, C., Charlot, S., Blaizot, J., & Brinchmann, J. 2012, MNRAS, 421, 2002
Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, ApJ, 559, 620
Peebles, P. 1980, The Large-scale structure of the Universe (Princeton University Press)
Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419
Pforr, J., Maraston, C., & Tonini, C. 2012, MNRAS, 422, 3285
Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, ApJ, 612, 168
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Press, W. H. & Schechter, P. 1974, ApJ, 187, 425
Prugniel, P. & Soubiran, C. 2001, A&A, 369, 1048
Prugniel, P. & Soubiran, C. 2004, ArXiv Astrophysics e-prints
Prugniel, P., Soubiran, C., Koleva, M., & Le Borgne, D. 2007, ArXiv Astrophysics e-prints
Rayner, J. T., Cushing, M. C., & Vacca, W. D. 2009, ApJS, 185, 289
Salpeter, E. E. 1955, ApJ, 121, 161
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Sandage, A. 1986, A&A, 161, 89
Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, A&AS, 96, 269

Schiavon, R. P. 2007, ApJS, 171, 146
Schiavon, R. P., Barbuy, B., Rossi, S. C. F., & Milone, A. 1997a, ApJ, 479, 902
Schiavon, R. P., Barbuy, B., & Singh, P. D. 1997b, ApJ, 484, 499
Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41
Shapley, A. E., Steidel, C. C., Erb, D. K., et al. 2005, ApJ, 626, 698
Smith, R. J., Lucey, J. R., & Conroy, C. 2015, MNRAS, 449, 3441
Spiniello, C., Trager, S. C., Koopmans, L. V. E., & Chen, Y. P. 2012, ApJL, 753, L32
Spinrad, H. & Taylor, B. J. 1971, ApJS, 22, 445
Tinsley, B. M. 1968, ApJ, 151, 547
Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, ApJ, 709, 1195
Valdes, F., Gupta, R., Rose, J. A., Singh, H. P., & Bell, D. J. 2004, ApJS, 152, 251
van Dokkum, P., Conroy, C., Villaume, A., Brodie, J., & Romanowsky, A. 2016, ArXiv e-prints
Vazdekis, A., Coelho, P., Cassisi, S., et al. 2015, MNRAS, 449, 1177
Villaume, A., Conroy, C., & Johnson, B. D. 2015, ApJ, 806, 82
Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, Ap&SS, 331, 1
Wing, R. F. & Ford, Jr., W. K. 1969, PASP, 81, 527
Worthey, G., Faber, S. M., Gonzalez, J. J., & Burstein, D. 1994, ApJS, 94, 687
Young, L. M., Bureau, M., Davis, T. A., et al. 2011, MNRAS, 414, 940

Chapter 2

A hierarchical Bayesian approach for reconstructing the Initial Mass Function of Single Stellar Populations

— **M. Dries, S.C. Trager and L.V.E. Koopmans** —

MNRAS, 2016, 463, 886

Abstract

Recent studies based on the integrated light of distant galaxies suggest that the initial mass function (IMF) might not be universal. Variations of the IMF with galaxy type and/or formation time may have important consequences for our understanding of galaxy evolution. We have developed a new stellar population synthesis (SPS) code specifically designed to reconstruct the IMF. We implement a novel approach combining regularization with hierarchical Bayesian inference. Within this approach we use a parametrized IMF prior to regulate a direct inference of the IMF. This direct inference gives more freedom to the IMF and allows the model to deviate from parametrized models when demanded by the data. We use Markov Chain Monte Carlo sampling techniques to reconstruct the best parameters for the IMF prior, the age, and the metallicity of a single stellar population. We present our code and apply our model to a number of mock single stellar populations with different ages, metallicities, and IMFs. When systematic uncertainties are not significant, we are able to reconstruct the input parameters that were used to create the mock populations. Our results show that if systematic uncertainties do play a role, this may introduce a bias on the results. Therefore, it is important to objectively compare different ingredients of SPS models. Through its Bayesian framework, our model is well-suited for this.

2.1 Introduction

Since its introduction by Salpeter (1955), the initial mass function (IMF) has been a key parameter in the study of stars, stellar populations and galaxy evolution. Salpeter initially parametrized the IMF as a single power law. However, it was recognized later on that when the IMF was extended down to the lowest stellar masses that it did not follow a single power law. Instead, a lognormal distribution (Miller & Scalo 1979), a multicomponent power law (Kroupa et al. 1993) or a combination of a lognormal distribution for low masses and a power law for higher masses (Chabrier 2003) were proposed as alternatives. Dabringhausen et al. (2008), among others, have shown that the latter two are in fact very similar.

Measurements of the IMF have long been based on direct star counts and mass estimates of resolved stars. These kinds of measurements are not possible for stars beyond the Local Group. Therefore, for many astrophysical studies the IMF has been assumed to be universal and similar to the one of the Milky Way. However, recent studies (Davé 2008; van Dokkum 2008; Treu et al. 2010; Graves & Faber 2010; Conroy & van Dokkum 2012; Cappellari et al. 2012; Spiniello et al. 2012, 2014; Ferreras et al. 2013; La Barbera et al. 2013) suggest that the IMF might not be universal on a cosmological scale, indicating that the relative number of low-mass stars in the population changes as a function of galaxy mass or velocity dispersion. This may have important consequences for the many properties of galaxies that are derived on the basis of the IMF, such as their stellar content, chemical enrichment history and even their evolutionary history: see e.g. Tinsley (1972).

Starting with Tinsley (1968), stellar population synthesis (SPS) models have been developed to transform the observable properties of a galaxy into a set of physical properties. Among the physical properties encrypted in the spectrum of a galaxy are its star formation history (SFH), the amount of gas and dust that it contains, its chemical composition and its IMF. However, deriving the low-mass end of the IMF on the basis of the spectrum of a galaxy is not straightforward. Dwarfs with masses $M < 0.4$ $M_\odot$ contribute only $\sim$1% to the integrated light of an old stellar population (Conroy & van Dokkum 2012). Nevertheless, they contribute 12 and 42% of the total stellar mass for a standard Kroupa IMF (Kroupa 2001) and a Salpeter IMF with a low-mass boundary of 0.1 $M_\odot$ and a high-mass boundary of 100 $M_\odot$, respectively. For younger stellar populations, the relative contribution of

low-mass stars to the spectrum is even less. In old stellar populations, the spectral similarity of low-mass stars and the most luminous stars (the K and M giants) further complicates the situation. However, a number of (gravity-sensitive) spectral features are known to be sensitive to either dwarfs or giants (Faber & French 1980; Schiavon et al. 1997a; Wing & Ford 1969; Schiavon et al. 1997b; Gorgas et al. 1993; Worthey et al. 1994; Schiavon 2007; Cenarro et al. 2003; Spiniello et al. 2012). The challenge for a SPS model is to extract this information from a spectrum.

Most SPS models are built upon three basic ingredients: a stellar evolution model in the form of isochrones as a function of age and metallicity, a stellar library, and an IMF. These ingredients form the basis of what is known as a single stellar population (SSP): a single, coeval population of stars with the same metallicity. The isochrone describes which stars are present in a stellar population, the stellar library provides a set of stellar spectra, and the IMF determines the distribution of stars along the isochrone. All of these ingredients have their own uncertainties. Models of stellar evolution are often one-dimensional codes and the results of these codes depend on the adopted prescriptions for uncertain factors, such as overshooting, rotation, interaction between binary stars, and mass loss. Stellar libraries may be theoretical, empirical or a combination of both. Both empirical and theoretical libraries have their own advantages and disadvantages. The assumption of a universal IMF is another source of uncertainty.

Real galaxies are not SSPs. Combining a set of SSPs with a SFH, a model for chemical evolution and possibly a dust model allows the construction of composite stellar populations (CSPs). To date, many different SPS models have been developed, e.g. Bruzual & Charlot (2003), Le Borgne et al. (2004), Maraston (2005), Conroy & van Dokkum (2012), and Vazdekis et al. (2012). Most of these models allow the user to change the IMF. Once an IMF or a set of different IMFs is defined, this allows the model to create synthesized spectra for a grid of different model parameters (including the IMF). The synthesized spectra are then compared with observed galaxy spectra to obtain values of, e.g., metallicity or IMF slope. Determining the best-fitting parameters is often done through a minimization technique, such as χ^2 minimization in for example Koleva et al. (2009). However, the ultimate goal of a SPS model would be a direct inference of the physical parameters from the spectrum.

Each SPS model uses its own set of ingredients, and the way in which these ingredients are combined also varies. This requires an objective manner to compare different SPS models with each other. A solution to this problem is provided by Bayesian inference. In this paper we develop a hierarchical Bayesian framework for SPS. Within our model, a parametrization of the IMF is used to construct a (flexible) IMF prior. Given this prior, our model allows for a direct inference of the piecewise IMF from the spectrum of an SSP. The outline of the paper is as follows. In Section 2 we discuss the Bayesian framework of our model. In Section 3 we describe how we construct a representative set of stellar templates as an input for our model. We then test our model by applying it to respectively mock SSPs and SSPs created by other SPS models in Sections 4 and 5.

2.2 Hierarchical Bayesian inference

Within a hierarchical Bayesian model there are multiple levels of inference. In this paper, we have two levels. The first level of inference assumes that a certain model family $\mathcal{H}$ can provide a proper description of the truth and tries to obtain the best-fit for the free parameters within that model family (parameter estimation). The second level of inference allows us to compare a set of different model families $\{\mathcal{H}_i\}$ and tries to infer the most probable model family given the data (model comparison). In analogy to the analysis presented by MacKay (1992), we derive a hierarchical Bayesian framework for modelling spectral energy distributions.

Neglecting the effect of extinction, which we will include in a future publication, the spectral energy distribution of a stellar population may be considered as the sum of the spectra of all the stars that it contains. This allows us to write the spectrum of the stellar population as a linear combination of a certain set of stellar templates. For an SSP, the stellar types that are present in the population are defined by an isochrone. The most important parameters that define an isochrone are its age t and metallicity [M/H]. An isochrone provides us with the stellar parameters (effective temperatures, surface gravities, luminosities, colors, initial masses, and current masses) of all the stellar types present in the corresponding SSP. These parameters are typically combined with a stellar library and an interpolator to create a spectrum $\boldsymbol{s}$ for each of the isochrone stars. This procedure is discussed in more detail in Section 2.3.

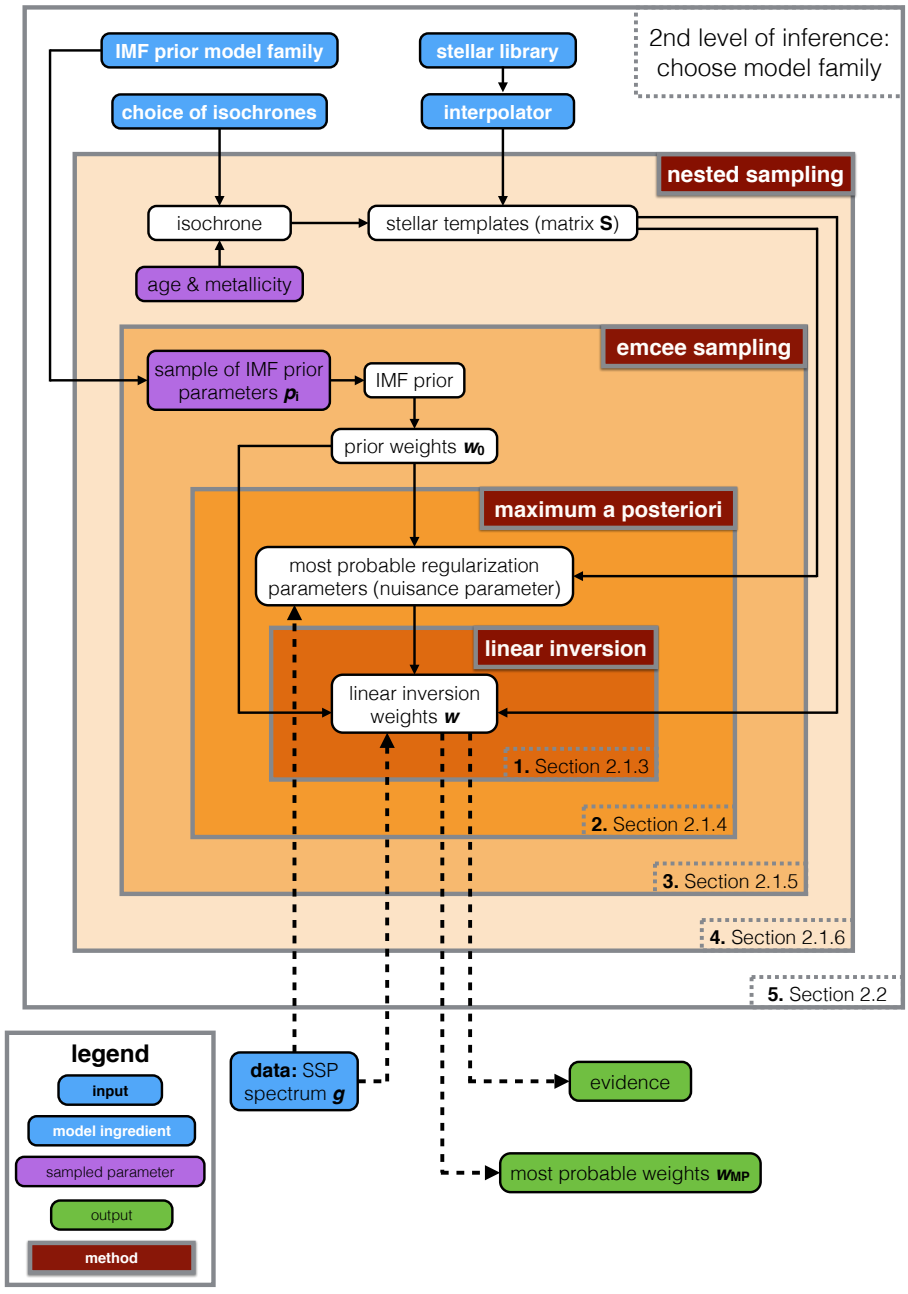

Figure 2.1 – Caption next page.

Figure 2.1 – Flow diagram illustrating the hierarchical nature of our model for an SSP. We have a spectrum which forms the input data of the model. At the outermost level, we define a model family $\mathcal{H}$ by choosing a set of isochrones, a stellar library, an interpolator, a regularization scheme, and a parametrization for the IMF prior. One level below, an age and metallicity define an isochrone. This isochrone is combined with the stellar library and the interpolator to create a set of stellar templates $\mathbf{S}$. The most probable age and metallicity are derived by calculating the evidence for every combination in a predefined age-metallicity grid. Going another level further down, a particular sample $p_{i,0}$ of the IMF prior model parameters p_i is transformed into a prior on the weights $\boldsymbol{w}_0$. The IMF prior model parameters are sampled using `emcee` (Foreman-Mackey et al. 2013). At the next level, the most probable value of the regularization parameter $\hat{\lambda}$ is determined given the data $\boldsymbol{g}$, the stellar templates $\mathbf{S}$ and the prior on the weights $\boldsymbol{w}_0$. Finally, at the innermost level the data $\boldsymbol{g}$, the stellar templates $\mathbf{S}$, the prior on the weights $\boldsymbol{w}_0$ and the most probable regularization parameter $\hat{\lambda}$ are combined to reconstruct the most probable weights w_{MP} and calculate the evidence for that particular set of parameters. At the highest level, we also calculate the evidence for a model family by marginalizing over all the free parameters in that model family. This allows us to compare different model families with each other and is referred to as the second level of inference.

Suppose that $\mathbf{S}$ is a matrix with the spectra of all the isochrone stars in its columns, such that $\mathbf{S}_{ij}$ corresponds to the i-th flux density bin $\boldsymbol{s}_i$ of the spectrum of isochrone star j. Since the isochrone is defined by its age and metallicity, $\mathbf{S} = \mathbf{S}(t, [\mathrm{M/H}])$ is implicitly also a function of age and metallicity (i.e. the age and metallicity define the isochrone, the isochrone defines a set of stars and their parameters which in turn are used to create a corresponding set of stellar spectra that goes into $\mathbf{S}$). Although here we consider SSPs, $\mathbf{S}$ might equally well contain the spectra of the stellar templates of a CSP. In that case $\mathbf{S}$ also becomes a function of the SFH of the stellar population.

If $\boldsymbol{w}$ is a vector with the number of stars for each stellar template in $\mathbf{S}$, the spectrum $\boldsymbol{g}$ of the stellar population is given by:

$$\boldsymbol{g} = \mathbf{S}\,\boldsymbol{w}. \tag{2.1}$$

For each star, an isochrone provides in general both the initial mass and the current mass (taking into account a prescription for possible mass loss). Since $\boldsymbol{w}$, hereafter called weights, represents the number of stars for each stellar template, the initial masses of the isochrone allow us to relate $\boldsymbol{w}$ to the IMF of the stellar population and vice versa. The IMF,

$$\xi(M) \equiv \frac{\mathrm{d}N}{\mathrm{d}M}, \tag{2.2}$$

of an SSP is related to $\boldsymbol{w}$ through

$$\xi(m_j) = \frac{\boldsymbol{w}_j}{m_{\rm high} - m_{\rm low}}, \tag{2.3}$$

where $\boldsymbol{w}_j$ is the number of stars of template j in the stellar population and m_j is the initial (rank-ordered) mass associated with template j by the isochrone. The boundaries $m_{\rm low}$ and $m_{\rm high}$ of the mass bin are defined such that

$$\begin{aligned} m_{\rm low} &= \frac{m_{j-1} + m_j}{2} \\ m_{\rm high} &= \frac{m_j + m_{j+1}}{2}. \end{aligned} \tag{2.4}$$

For the lowest mass template, $m_{\rm low} = m_{\rm LMCO}$ (low-mass-cut-off of the IMF) and for the highest mass template we take $m_{\rm high} = m_j$. In this way, $\boldsymbol{w}_j$ corresponds to the number of stars in the mass bin $(m_{\rm low}, m_{\rm high})$. The way in which $m_{\rm low}$ and $m_{\rm high}$ are defined ensures that mass bins never overlap.

In this section we first discuss how to find the most probable distribution of weights $\boldsymbol{w}_{\rm MP}$ by using the combination of regularization and hierarchical Bayesian inference. Subsequently we discuss how Markov Chain Monte Carlo techniques may be used to reconstruct the parameters of a certain IMF prior parametrization and to find the age and metallicity of the SSP. As a last step, we show how different model families may be compared on the basis of their Bayesian evidence. The hierarchical nature of the different steps in the model is illustrated in Fig. 2.1.

2.2.1 The first level of inference

At the first level of inference, we assume that model family $\mathcal{H}$ is the correct model family and we try to infer the model parameters given the data $\boldsymbol{g}$. A model family $\mathcal{H}$ is defined on the one hand by the set of stellar templates $\mathbf{S}$ (e.g. a set of SPS templates as a function of age and metallicity) and on the other hand by the parametrization of the IMF prior, which defines the space of possible priors on the weights $\boldsymbol{w}$. In Section 2.3 we discuss how to construct a representative set of stellar templates whereas the parametrization of the IMF prior, and hence the prior on the weights, is discussed in more detail in Section 2.2.1.2.

Given a certain model family $\mathcal{H}$, we infer the number of stars $\boldsymbol{w}$ for each of the templates in our model given the spectrum of a stellar population $\boldsymbol{g}$.

Note that in this paper we consider SSPs but in principle this can also be done for CSPs if we include the SFH and chemical evolution of the stellar population in our model as well, as we plan to do in the future.

2.2.1.1 The most likely solution

In reality the spectrum of a stellar population contains noise, such that the observed spectrum of the stellar population becomes

$$\boldsymbol{g} = \mathbf{S}\,\boldsymbol{w} + \boldsymbol{n}, \tag{2.5}$$

in which $\boldsymbol{n}$ represents the noise in the data. Assuming that the noise is Gaussian distributed, the likelihood of the data $\mathcal{L}(\boldsymbol{g}|\boldsymbol{w},\mathbf{S})$ given the weights $\boldsymbol{w}$ and the stellar templates $\mathbf{S}$ is

$$\mathcal{L}(\boldsymbol{g}|\boldsymbol{w},\mathbf{S}) = \frac{\exp[-E_{_D}(\boldsymbol{g}|\boldsymbol{w},\mathbf{S})]}{Z_{_D}}, \tag{2.6}$$

in which

$$\begin{aligned} E_{_D}(\boldsymbol{g}|\boldsymbol{w},\mathbf{S}) &= \frac{1}{2}(\mathbf{S}\,\boldsymbol{w}-\boldsymbol{g})^{\mathrm{T}}\mathbf{C}_{\mathrm{D}}^{-1}(\mathbf{S}\,\boldsymbol{w}-\boldsymbol{g}) \\ &\equiv \frac{1}{2}\chi^2, \end{aligned} \tag{2.7}$$

where $\mathbf{C}_{\mathrm{D}}$ is the covariance matrix. The likelihood is normalized by

$$Z_{_D} = (2\pi)^{N_{_D}/2}(\det \mathbf{C}_{\mathrm{D}})^{1/2}, \tag{2.8}$$

in which $N_{_D}$ is the number of data points in the spectrum.

The most likely solution $\boldsymbol{w}_{\mathrm{ML}}$ may be found by maximizing the likelihood function $\mathcal{L}(\boldsymbol{g}|\boldsymbol{w},\mathcal{H}_i)$ defined in equation 2.6. Maximizing the likelihood implies minimizing $E_{_D}$, so that we obtain

$$\nabla E_{_D}(\boldsymbol{w}_{\mathrm{ML}}) \equiv \frac{\partial E_{_D}(\boldsymbol{w}_{\mathrm{ML}})}{\partial \boldsymbol{w}} = 0. \tag{2.9}$$

The solution to this equation is given by

$$\boldsymbol{w}_{\mathrm{ML}} = (\mathbf{S}^{\mathrm{T}}\mathbf{C}_{\mathrm{D}}^{-1}\mathbf{S})^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{C}_{\mathrm{D}}^{-1}\boldsymbol{g}. \tag{2.10}$$

Finding $\boldsymbol{w}_{\mathrm{ML}}$ is in general an ill-posed problem. Therefore we use a prior on the weights $\boldsymbol{w}$ to regularize the solution that we obtain and to find the most probable distribution of weights $\boldsymbol{w}_{\mathrm{MP}}$.

2.2.1.2 The prior

Suppose that within a model family $\mathcal{H}$, the IMF prior is parametrized by a set of (non-linear) parameters which we call p_i. For example the IMF prior may be parametrized as a power law which is defined by its slope α and the normalization $C_{\rm norm}$: in that case $p_i = \{\alpha, C_{\rm norm}\}$. If we take one particular combination $p_{i,0}$ of the parameters p_i, this completely defines a prior $\xi_0(p_{i,0}, M)$ on the IMF. By using the initial masses associated to the templates through the isochrone, the prior ξ_0 on the IMF translates into a prior on the weights which we refer to as $\boldsymbol{w}_0$. Since $\xi(M) \equiv \dfrac{\mathrm{d}N}{\mathrm{d}M}$, the number of stars that we have for template j is given by

$$w_{0,j} = \int_{m_{\rm low}}^{m_{\rm high}} \xi_0(p_{i,0}, M)\mathrm{d}M, \tag{2.11}$$

where $m_{\rm low}$ and $m_{\rm high}$ are defined in equation 2.4. So within one model family $\mathcal{H}$, there is a range of different models $\mathcal{H}_0$ that are defined by different priors $\boldsymbol{w}_0$. The allowed range of priors $\boldsymbol{w}_0$ within the model family is defined by the functional form of the IMF prior and its parameters p_i. Note that that the latter may have their own priors as well.

Once we have transformed the prior on the IMF ξ_0 into a prior on the weights $\boldsymbol{w}_0$, we define the regularization function $E_S(\boldsymbol{w}|\boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1})$ as

$$E_S(\boldsymbol{w}|\boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1}) = \frac{1}{2}(\boldsymbol{w} - \boldsymbol{w}_0)^{\rm T}\mathbf{C}_{\rm pr}^{-1}(\boldsymbol{w} - \boldsymbol{w}_0), \tag{2.12}$$

where $\mathbf{C}_{\rm pr}^{-1} = \nabla\nabla E_S(\boldsymbol{w})$ is the (constant) Hessian of E_S. Hence the regularization function puts a penalty on $\boldsymbol{w}$ for deviating from the prior distribution of weights $\boldsymbol{w}_0$. Deviations are only possible if the data require it (i.e. if a deviation increases the likelihood more than it decreases the prior). Note that $\mathbf{C}_{\rm pr}^{-1}$ is part of $\mathcal{H}$ as it is a model-dependent choice that relates to the form of regularization being used: we might for example use the identity matrix or enforce smoothness. Given the regularization function, the prior probability function may be expressed as

$$\Pr(\boldsymbol{w}|\lambda, \boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1}) = \frac{\exp[-\lambda E_S(\boldsymbol{w}|\boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1})]}{Z_S(\lambda)}, \tag{2.13}$$

where λ is the regularization parameter and the prior probability function is normalized by Z_S. A larger regularization parameter implies that there

is more emphasis on the prior and less on the likelihood. In Section 2.2.1.5 we show how the value of the regularization parameter may be derived in a Bayesian manner. The regularization parameter is therefore a nuisance parameter that must be marginalized over in the results.

2.2.1.3 The posterior

We have seen that a model family $\mathcal{H}$ is defined by the stellar templates $\mathbf{S}$, the parametrization of the IMF p_i and by the choice of the Hessian $\mathbf{C}_{\rm pr}^{-1}$, such that $\mathcal{H} = \{\mathbf{S}, p_i, \mathbf{C}_{\rm pr}^{-1}\}$. Within such a model family there exists a range of models $\mathcal{H}_0 = \{\mathbf{S}, \boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1}\}$, in which $\boldsymbol{w}_0$ is related to one particular choice $p_{i,0}$ of the IMF prior parametrization. In this way, each model $\mathcal{H}_0$ is defined by a different prior $\boldsymbol{w}_0$. Hence the prior $\boldsymbol{w}_0$ is not fixed but should be considered as a flexible entity that is allowed to change within the boundaries of the IMF parametrization p_i. For every model $\mathcal{H}_0$, the likelihood and the prior $\boldsymbol{w}_0$ are combined to find the most probable distribution of weights $\boldsymbol{w}_{\rm MP}$. Defining $M(\boldsymbol{w}|\mathcal{H}_0)$ as

$$M(\boldsymbol{w}|\mathcal{H}_0) = E_{\scriptscriptstyle D}(\boldsymbol{w}|\mathbf{S}) + \lambda E_{\scriptscriptstyle S}(\boldsymbol{w}|\boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1}), \tag{2.14}$$

we apply Bayes' theorem to combine the likelihood function and the prior probability function into the posterior probability function

$$\begin{aligned} P(\boldsymbol{w}|\boldsymbol{g}, \lambda, \mathcal{H}_0) &= \frac{\mathcal{L}(\boldsymbol{g}|\boldsymbol{w}, \mathbf{S}) \cdot \Pr(\boldsymbol{w}|\lambda, \boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1})}{P(\boldsymbol{g}|\lambda, \mathcal{H}_0)} \\ &= \frac{\exp[-M(\boldsymbol{w})]}{Z_{\scriptscriptstyle M}(\lambda)}, \end{aligned} \tag{2.15}$$

where the posterior is normalized by $Z_{\scriptscriptstyle M}(\lambda)$. The last equation shows that the posterior probability distribution for the weights $\boldsymbol{w}$ of the stellar templates is controlled by two functions. On the one hand there is the 'goodness of fit' represented by $E_{\scriptscriptstyle D}(\boldsymbol{w}|\mathbf{S})$ and on the other hand there is the deviation of the weights from the prior represented by the regularization function $E_{\scriptscriptstyle S}(\boldsymbol{w}|\boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1})$. The balance between these two functions is set by the regularization parameter[1].

[1]Naively one might think $\lambda = 0$ will give the highest posterior, but since the prior is normalized, lowering λ makes the width of the prior very large, hence lowering the probability density at the position where the likelihood peaks. This lowers the posterior probability. Making λ larger will increase the latter, but might make the fit to the data more difficult, lowering the likelihood. Balancing these is the Bayesian equivalent to Occam's razor, finding the simplest model that fits the data.

To find the most probable solution $\boldsymbol{w}_{\rm MP}$, we have to maximize the posterior probability density function (Equation 2.15). Maximizing $P(\boldsymbol{w}|\boldsymbol{g},\lambda,\mathcal{H}_0)$ implies minimizing $M(\boldsymbol{w}|\mathcal{H}_0)$, so that we have $\nabla M(\boldsymbol{w}_{\rm MP}) = 0$. Defining $\mathbf{B} \equiv \nabla\nabla E_D(\boldsymbol{w}) = \mathbf{S}^{\rm T}\mathbf{C}_{\rm D}^{-1}\mathbf{S}$ as the Hessian of E_D and using the definition of E_S from equation 2.12, we have for the most probable solution

$$\boldsymbol{w}_{\rm MP} = \mathbf{A}^{-1}(\mathbf{S}^{\rm T}\mathbf{C}_{\rm D}^{-1}\boldsymbol{g} + \lambda\mathbf{C}_{\rm pr}^{-1}\boldsymbol{w}_0), \tag{2.16}$$

where $\mathbf{A} \equiv \nabla\nabla M(\boldsymbol{w}) = \mathbf{B} + \lambda\mathbf{C}_{\rm pr}^{-1}$ is the Hessian of $M(\boldsymbol{w})$. In practice we solve equation 2.16 by using non-negative least squares[2] (NNLS) to ensure a physically meaningful solution (i.e. the number of stars cannot be negative: $\boldsymbol{w} \not< 0$).

The most probable solution depends on the model $\mathcal{H}_0 = \{\mathbf{S}, \boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1}\}$ as well as on the regularization parameter λ that regulates the balance between the 'goodness of fit' and the penalty term resulting from the regularization function. The inversion of the most probable weights is represented by the inner block in Fig. 2.1. To find $\boldsymbol{w}_{\rm MP}$, the inner block needs information from the outer levels: a set of stellar templates, a prior on the weights and a regularization parameter. In Section 2.2.1.5 we show how to find the most probable value of the regularization parameter given the model and the data.

2.2.1.4 Uncertainties of the most probable weights

Using a second order Taylor expansion for $M(\boldsymbol{w})$ around $\boldsymbol{w}_{\rm MP}$, we may approximate $M(\boldsymbol{w})$ as

$$M(\boldsymbol{w}|\mathcal{H}_0) = M(\boldsymbol{w}_{\rm MP}) + \frac{1}{2}\Delta\boldsymbol{w}^{\rm T}\mathbf{A}\Delta\boldsymbol{w}, \tag{2.17}$$

with $\Delta\boldsymbol{w} = \boldsymbol{w} - \boldsymbol{w}_{\rm MP}$. This allows us to approximate the posterior as

$$P(\boldsymbol{w}|\boldsymbol{g},\lambda,\mathcal{H}_0) \approx P(\boldsymbol{w}_{\rm MP}) \cdot \exp\left[-\frac{1}{2}\Delta\boldsymbol{w}^{\rm T}\mathbf{A}\Delta\boldsymbol{w}\right]. \tag{2.18}$$

From this equation we see that the posterior may be approximated locally as a multivariate Gaussian distribution with covariance matrix $\mathbf{A}^{-1}$. The marginalized errors on the individual weights $\boldsymbol{w}_{\rm MP}$ resulting from the linear inversion may be obtained by taking the square root of the diagonal elements in $\mathbf{A}^{-1}$.

[2] Lawson & Hanson (1995)

2.2.1.5 The regularization parameter

To find the optimal regularization parameter $\hat{\lambda}$, we have to find the maximum value for the probability density function $P(\lambda|\boldsymbol{g}, \mathcal{H}_0)$. According to Bayes' theorem $P(\lambda|\boldsymbol{g}, \mathcal{H}_0)$ is written as

$$P(\lambda|\boldsymbol{g}, \mathcal{H}_0) = \frac{P(\boldsymbol{g}|\lambda, \mathcal{H}_0) \cdot P(\lambda)}{P(\boldsymbol{g}|\mathcal{H}_0)} \propto P(\boldsymbol{g}|\lambda, \mathcal{H}_0) \cdot P(\lambda). \tag{2.19}$$

Neglecting the normalization constant $P(\boldsymbol{g}|\mathcal{H}_0)$, the function to consider for optimizing λ is the product of the likelihood $P(\boldsymbol{g}|\lambda, \mathcal{H}_0)$ and the prior $P(\lambda)$. Note that the likelihood term $P(\boldsymbol{g}|\lambda, \mathcal{H}_0)$ appears as the normalizing constant of equation 2.15: this term is often referred to as the evidence. Using equations 2.6, 2.8, and 2.13-2.15 we have

$$P(\boldsymbol{g}|\lambda, \mathcal{H}_0) = \frac{Z_{\scriptscriptstyle M}(\lambda)}{Z_{\scriptscriptstyle D} \cdot Z_{\scriptscriptstyle S}(\lambda)}. \tag{2.20}$$

Using the definition of $E_{\scriptscriptstyle S}$ from equation 2.12, the normalization of the prior becomes

$$\begin{aligned} Z_{\scriptscriptstyle S} &= \int \mathrm{d}^{N_\mathrm{w}} \boldsymbol{w} \cdot \exp(-\lambda E_{\scriptscriptstyle S}) \\ &= \left(\frac{2\pi}{\lambda}\right)^{N_\mathrm{w}/2} (\det \mathbf{C}_\mathrm{pr}^{-1})^{-1/2}, \end{aligned} \tag{2.21}$$

where N_w is the number of stellar templates in model $\mathcal{H}_0$. Using the Taylor expansion from equation 2.17 we have for $Z_{\scriptscriptstyle M}$

$$\begin{aligned} Z_{\scriptscriptstyle M}(\lambda) &= \int \mathrm{d}^{N_\mathrm{w}} \boldsymbol{w} \cdot \exp(-M(\boldsymbol{w})) \\ &= e^{-M(\boldsymbol{w}_\mathrm{MP})} (2\pi)^{N_\mathrm{w}/2} (\det \mathbf{A})^{-\frac{1}{2}}. \end{aligned} \tag{2.22}$$

Combining equations 2.8 and 2.20-2.22 allows us to write the logarithm of $P(\boldsymbol{g}|\lambda, \mathcal{H}_0)$ as

$$\begin{aligned} \log P(\boldsymbol{g}|\lambda, \mathcal{H}_0) = &-M(\boldsymbol{w}_\mathrm{MP}) - \frac{1}{2}\log(\det \mathbf{A}) \\ &+ \frac{N_\mathrm{w}}{2}\log\lambda + \frac{1}{2}\log(\det \mathbf{C}_\mathrm{pr}^{-1}) \\ &- \frac{N_\mathrm{d}}{2}\log 2\pi + \frac{1}{2}\log(\det \mathbf{C}_\mathrm{D}^{-1}). \end{aligned} \tag{2.23}$$

Since we do not know a priori the value of λ nor its order of magnitude, we choose a flat prior in $\log\lambda$ such that $P(\lambda) \propto 1/\lambda$. The optimal regularization parameter is then found by solving $\dfrac{d}{d\log\lambda}\log\left(P(\boldsymbol{g}|\lambda,\mathcal{H}_0)\cdot P(\lambda)\right) = 0$, which results in the following non-linear expression for the most probable value of the regularization parameter $\hat{\lambda}$

$$\hat{\lambda}E_S(\boldsymbol{w}_{\rm MP}) = \frac{N_{\rm w}}{2} - \frac{1}{2}\hat{\lambda}\mathrm{Tr}(\mathbf{A}^{-1}\mathbf{C}_{\rm pr}^{-1}) - 1, \tag{2.24}$$

where the last term in this equation originates from the prior on λ. This equation may be solved by using a non-linear solver. The process of finding the most probable regularization parameter is represented by block 2 in Fig. 2.1. Note that for every step in finding the solution to equation 2.24, the model has to go to the inner block to find the most probable weights $\boldsymbol{w}_{\rm MP}$. Instead of solving for the most probable regularization parameter, λ can in principle also be sampled as a nuisance parameter together with the other non-linear parameters of the model.

2.2.1.6 Reconstructing the IMF model parameters

To reconstruct the parameters p_i of the IMF parametrization in a model family $\mathcal{H}$, we compare the different models $\mathcal{H}_0$ in that model family with each other. The posterior probability of a certain model $\mathcal{H}_0$ is given by

$$P(\mathcal{H}_0|\boldsymbol{g}) \propto P(\boldsymbol{g}|\mathcal{H}_0)\cdot P(\mathcal{H}_0). \tag{2.25}$$

In the case of a flat prior $P(\mathcal{H}_0)$, models may be compared on the basis of the likelihood term $P(\boldsymbol{g}|\mathcal{H}_0)$. Taking into account that $P(\boldsymbol{g}|\mathcal{H}_0)$ is actually a marginalization over λ, we may write it as

$$P(\boldsymbol{g}|\mathcal{H}_0) = \int P(\boldsymbol{g}|\lambda,\mathcal{H}_0)\cdot P(\lambda|\boldsymbol{g},\mathcal{H}_0)\mathrm{d}\lambda, \tag{2.26}$$

where $P(\boldsymbol{g}|\lambda,\mathcal{H}_0)$ is the evidence derived in equation 2.23.

If we make the assumption that $P(\lambda|\boldsymbol{g},\mathcal{H}_0)$ is a strongly peaked function at the most probable value $\hat{\lambda}$ (MacKay 1992), we may approximate it by a delta function centred on $\hat{\lambda}$ so that we obtain:

$$\begin{aligned} P(\boldsymbol{g}|\mathcal{H}_0) &= \int P(\boldsymbol{g}|\lambda,\mathcal{H}_0)\cdot P(\lambda|\boldsymbol{g},\mathcal{H}_0)\mathrm{d}\lambda \\ &= \int P(\boldsymbol{g}|\lambda,\mathcal{H}_0)\cdot \delta(\hat{\lambda})\mathrm{d}\lambda \\ &= P(\boldsymbol{g}|\hat{\lambda},\mathcal{H}_0), \end{aligned} \tag{2.27}$$

and we may rank the different models on the basis of $P(\boldsymbol{g}|\hat{\lambda}, \mathcal{H}_0)$, i.e. the evidence from equation 2.23 evaluated for the most probable regularization constant $\hat{\lambda}$. Note that if we compare two models $\mathcal{H}_{0,1}$ and $\mathcal{H}_{0,2}$ on the basis of their evidence, we are actually interested in the ratio of the evidence for model $\mathcal{H}_{0,1}$ by the evidence for model $\mathcal{H}_{0,2}$. This ratio is called the Bayes factor K. Values of $K > 10^{1/2}$ and $K > 10^1$ may be considered as, respectively, substantial and strong evidence in favour of $\mathcal{H}_{0,1}$ whereas $K > 10^2$ is in general considered as decisive evidence in favour of $\mathcal{H}_{0,1}$ (Jeffreys 1961).

The ability that we now have to quantify the posterior of a model $\mathcal{H}_0(p_i)$ on the basis of the evidence allows us to use Monte Carlo sampling techniques to reconstruct the posterior probability distribution of the non-linear IMF prior model parameters. In this paper we use `emcee` (Foreman-Mackey et al. 2013) to sample the IMF prior parameters p_i. The reconstruction of the IMF prior model parameters is visualized by block 3 in Fig. 2.1. For every sample of the IMF prior model parameters $p_{i,0}$, the model constructs a corresponding prior $\boldsymbol{w}_0$. Then the model finds the most probable regularization parameter at the level below. Finally the model determines the most probable weights and calculates the evidence which may in turn be used to compare different samples $p_{i,0}$ of the IMF prior model parameters.

2.2.1.7 The age and metallicity of the SSP

Before we can actually sample the IMF prior model parameters p_i, we have to define the set of stellar templates that we are going to use. For SSPs, the stellar templates are defined by an isochrone of a certain age and metallicity. If we want to compare a combination of two different ages and metallicities (i.e. $\mathbf{S}(t_1, [\mathrm{M/H}]_1)$ vs. $\mathbf{S}(t_2, [\mathrm{M/H}]_2)$) we may once again use the Bayes factor

$$K_{12} = \frac{P(\boldsymbol{g}|\mathbf{S}(t_1, [\mathrm{M/H}]_1))}{P(\boldsymbol{g}|\mathbf{S}(t_2, [\mathrm{M/H}]_2))}. \tag{2.28}$$

In this equation, $P(\boldsymbol{g}|\mathbf{S}(t_j, [\mathrm{M/H}]_j))$ is defined as

$$P(\boldsymbol{g}|\mathbf{S}(t_j, [\mathrm{M/H}]_j)) = \int P(\boldsymbol{g}|p_i, \mathbf{S}(t_j, [\mathrm{M/H}]_j)) \cdot \Pr(p_i) \mathrm{d}p_i, \tag{2.29}$$

which represents the evidence (or marginal likelihood) for the templates defined by $\{t_j, [\mathrm{M/H}]_j\}$ (i.e. in determining the evidence for the age and

metallicity combination $\{t_j, [\mathrm{M/H}]_j\}$ we marginalize over all parameters in the inner layers, among which the IMF prior model parameters p_i). To find the most probable age and metallicity, we determine the evidence for each entry in a predefined age-metallicity grid. The age-metallicity combination that results in the highest evidence is then used to refine the sampling of the IMF prior model parameters p_i with `emcee`. Note that we are only allowed to do this if there is a clear peak for the evidence in our age-metallicity grid. Otherwise, the resulting distributions for the IMF model parameters should be marginalized over all ages and metallicities.

An efficient method for determining the integral in equation 2.29 is provided by nested sampling (Skilling 2004). We calculate evidences by using `Multinest` (Feroz & Hobson 2008; Feroz et al. 2009, 2013). To implement the Monte Carlo sampling techniques that we use in this paper, we have developed our code as a pipeline of `cosmoSIS` (Zuntz et al. 2015). `CosmoSIS` is a cosmological parameter estimation code that brings together different inference tools, including `Multinest` and `emcee`.

The reconstruction of the (most probable) age and metallicity is represented by block 4 in Fig. 2.1. For every age and metallicity, the model selects an isochrone which is combined with the stellar library and the interpolator to create a set of stellar templates. At the level below, `Multinest` requires a complete sample of the IMF model parameters p_i which allows it to marginalize over these parameters and calculate the evidence for the corresponding age and metallicity. This step still belongs to the first level of inference as we are trying to determine a set of parameters (i.e. age and metallicity).

2.2.2 The second level of inference

For the first level of inference we assume a certain model family $\mathcal{H}$. This model family is defined by the choice of isochrones, the stellar library, the interpolator, the regularization method and the parametrization of the IMF. The choice we make in this work to restrict ourselves to SSPs is also a model-dependent choice. We define $p_{\mathcal{H}}$ as the set of parameters that defines a model family. The second level of inference allows us to compare different model families with each other. As an example, for a given dataset we might want to compare two different IMF prior parametrizations: e.g. a double power law parametrization versus a lognormal parametrization.

According to Bayes' theorem, the posterior of a model family $\mathcal{H}$ given the data $\boldsymbol{g}$ is

$$P(\mathcal{H}|\boldsymbol{g}) \propto P(\boldsymbol{g}|\mathcal{H}) \cdot P(\mathcal{H}). \tag{2.30}$$

Assuming a flat prior $P(\mathcal{H})$ over the model families, the posterior of a model family is proportional to the likelihood $P(\boldsymbol{g}|\mathcal{H})$. This likelihood term is a marginalization over the free parameters of the model family

$$P(\boldsymbol{g}|\mathcal{H}) = \int P(\boldsymbol{g}|p_{\mathcal{H}}, \mathcal{H}) \cdot P(p_{\mathcal{H}}|\mathcal{H}) \mathrm{d}p_{\mathcal{H}}. \tag{2.31}$$

This integral may be determined by using e.g. `Multinest` and gives us the evidence for a certain model family.

If we return to the example where we want to compare a double power law parametrization of the IMF with a lognormal parametrization, the relevant model parameters $P_{\mathcal{H}}$ to marginalize over are the parameters of the IMF prior parametrization p_i. This is however only true if we find a sufficiently strong peak in the evidence of the age-metallicity grid that justifies the use of an SSP. If this is not the case, we should in principle also marginalize over all ages and metallicities to obtain the evidence for a model family. In the current paper, since we test only SSP models, there is no strong need to do this, but we plan to further expand the code to sample directly over the space of age and metallicity and then expand the code to enable modelling CSPs as well. The second level of inference in our model is represented by the outer shell in Fig. 2.1.

Now that we have discussed the general setup of our model, in the next section we discuss the particular set of ingredients that we use to apply our model in this paper.

2.3 Stellar templates

In this section we describe the basic ingredients that we use to create the stellar templates as an input for our model and that form the columns of the matrix $\mathbf{S}$. Here we consider the example of an SSP as a function of age and metallicity. Therefore the age and metallicity are two free parameters of the model, although they are here solved on a regular grid of values and therefore are not part of the 'continuous' set of parameters $\boldsymbol{w}$, λ and the IMF prior model parameters p_i. Those continuous parameters can be

inferred and the evidence obtained for each chosen age and metallicity via nested sampling as discussed in Section 2.2.1.7. Since that evidence is the probability density for a chosen age and metallicity, it can be used for model comparison.

Note that although we describe one particular set of ingredients in this section, in principle these ingredients may be substituted by any other set of ingredients. The approach described in Section 2.2 will still be valid, as long as we are able to construct a representative set of stellar templates.

2.3.1 Isochrones

The stars that are present in an SSP are defined by an isochrone. For a given age and metallicity, an isochrone provides us with the effective temperatures, surface gravities, masses and luminosities of the stars in an SSP corresponding to that particular age and metallicity.

Within our model we use the Padova isochrones described in Marigo et al. (2008). These isochrones may in principle be replaced with other models and different isochrone models may be assessed based on the evidence. The age and metallicity that define an isochrone are in principle continuous parameters. However, for every combination of age and metallicity we need to create an isochrone and for every isochrone star we need to interpolate a corresponding spectrum. Since isochrone determination and spectrum interpolation is time consuming, we create the stellar templates before running the model. Therefore we model age and metallicity as discrete parameters. We define a grid of ages and metallicities such that $\log \text{age} = \{8.0, 10.11\}$ and $[\text{M/H}] = \log(Z/Z_\odot) = \{-1.0, 0.4\}$ with $Z_\odot = 0.019$. The spacings of the grid are $\Delta \log t\,[\text{Gyr}] \approx 0.062$ and $\Delta[\text{M/H}] = 0.05$, respectively.

2.3.2 Stellar library

To construct a spectrum for a given set of stellar parameters, the starting point is a stellar library. Currently we use the (empirical) MILES stellar library (Sánchez-Blázquez et al. (2006)), consisting of approximately 1000 stars. Once again, note that this is only one particular choice, and in future work we plan to extend our model to include the X-Shooter Spectral Library (Chen et al. 2014). Although the MILES library covers a broad range in atmospheric parameters, empirical libraries have the disadvantage that they provide a limited coverage of the Hertzsprung-Russel (HR) diagram.

Fig. 2.2 shows the HR diagram of the MILES stars. The figure also shows four isochrones used in our model. One can see that, although in general the coverage of the isochrones is quite good, there are some regions of parameter space where there is clearly a lack of stars. Especially for the low-mass end and the upper giant and asymptotic giant branches, it is apparent that the stars in the library do not fully cover the parameter space defined by the isochrone stars. As a consequence, there can potentially be significant uncertainties in the stellar spectra that are constructed in these regions.

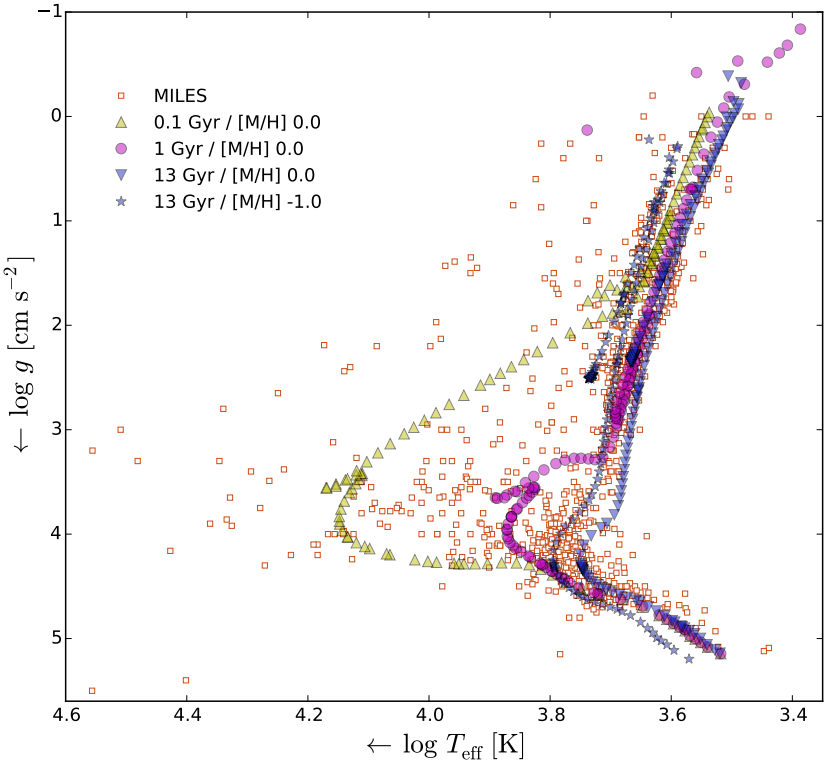

Figure 2.2 – Effective temperatures and surface gravities of stars in the MILES library (*red squares*). Also shown are three isochrones for solar metallicity with ages 0.1 Gyr (*yellow upper triangles*), 1 Gyr (*magenta circles*), and 13 Gyr (*blue lower triangles*). For the 13 Gyr isochrone, a low metallicity variant with [M/H] = −1.0 is also shown (*blue stars*).

2.3.3 Interpolator

The limited coverage of the HR diagram by empirical libraries requires a method to attach the stars in the library to the isochrones. We use an interpolator to do this, which for a given set of stellar parameters tries to interpolate between the surrounding spectra to create a representative stellar spectrum.

The idea behind such an interpolator is to create a function that interpolates between the spectra in the stellar library, allowing us to construct stellar spectra at all relevant locations in the HR diagram. Before creating an interpolator, one has to define the parameters that are required to model the spectrum of a star. In addition to the effective temperature and surface gravity, this would in principle require detailed knowledge of all chemical abundances in the star. However, the isochrones only define the overall metallicity [M/H], whereas the stars in the MILES library have measured values of [Fe/H] available. Therefore we choose to use an interpolator $S_\lambda(T_{\rm eff}, \log g, [{\rm Fe/H}])$ that interpolates the spectra of the stars in the three dimensional space of effective temperature, surface gravity and [Fe/H]. In addition, we assume that [M/H] = [Fe/H] to make the conversion from the isochrones to the stellar library straightforward. For the parameters of the stars in the MILES library, we use the values derived by Cenarro et al. (2007).

Interpolating between stellar spectra may be done by using either a local approach or a global approach. Within the local approach described in Vazdekis et al. (2003), the spectra in the library that surround the point for which we want to create a spectrum are weighted and combined to create a representative spectrum for that particular point. The global approach described in Prugniel et al. (2011) fits a polynomial to each of the spectral bins individually. This polynomial may then be used to determine the flux in each of the bins for the required set of atmospheric parameters.

In this work, we use a local approach very similar to that described in Vazdekis et al. (2003). Before we build the interpolator, we normalize the stars in the MILES library such that they have the same magnitude in the (Johnson) V-band. Suppose that we want to create a spectrum for the point $\{\theta_0, \log g_0, [{\rm Fe/H}]_0\}$ (where $\theta = 5040/T_{\rm eff}$). Within the three dimensional space of θ, $\log g$ and [Fe/H], this point is surrounded by eight cubes. The first step of the interpolator consists of finding the nearby stars that are present in each of these eight boxes. The initial size of each of

these cubes is $1.5\sigma_\theta \times 1.5\sigma_{\log g} \times 1.5\sigma_{[\mathrm{Fe/H}]}$, in which σ_p corresponds to the typical uncertainty in the parameters $p = \{\theta, \log g, [\mathrm{Fe/H}]\}$. These typical uncertainties are defined on the basis of the local density of stars ρ, such that

$$\sigma_p = \sigma_{p,\mathrm{m}} \cdot \exp\left(\left(\frac{\rho - \rho_\mathrm{M}}{\rho_\mathrm{M}}\right)^2 \ln \frac{\sigma_{p,\mathrm{M}}}{\sigma_{p,\mathrm{m}}}\right). \tag{2.32}$$

In this equation ρ_M is the maximum density of stars in the grid, which is taken as the 99.7 percentile of all densities in the grid. For the minimum uncertainty $\sigma_{p,\mathrm{m}}$ and maximum uncertainty $\sigma_{p,\mathrm{M}}$ we use the same values as Vazdekis et al. (2003): $\sigma_{\theta,\mathrm{m}} = 0.009$, $\sigma_{\theta,\mathrm{M}} = 0.17$, $\sigma_{\log g,\mathrm{m}} = 0.18$, $\sigma_{\log g,\mathrm{M}} = 0.51$, $\sigma_{[\mathrm{Fe/H}],\mathrm{m}} = 0.09$ and $\sigma_{[\mathrm{Fe/H}],\mathrm{M}} = 0.41$. As an additional constraint for σ_θ, σ_{T_eff} should lie within 60 K $\leq \sigma_{T_\mathrm{eff}} \leq$ 3350 K. If no stars are found in one of the boxes, the size of the box is enlarged in steps of $0.5\sigma_p$ along each of its axes until at least one star is found or the axes reach a size of $10\sigma_p$. Note that the metallicity parameter is only taken into account for stars with 4000 K $\leq T_\mathrm{eff} \leq$ 9000 K. Outside of this range the uncertainty in the metallicity is relatively large and in addition there is a significant number of stars with unknown metallicity.

As a next step, we create a representative spectrum for each of the boxes that contain stars. To create the spectrum of a box, each of its stars is assigned a weight W_s such that

$$\begin{aligned} W_\mathrm{s} = &\frac{\mathrm{SN}_\mathrm{s}^2}{\mathrm{SN}_\mathrm{max}^2} \cdot \exp\left(-\left(\frac{\theta_\mathrm{s} - \theta_0}{\sigma_\theta}\right)^2\right) \\ &\cdot \exp\left(-\left(\frac{\log g_\mathrm{s} - \log g_0}{\sigma_{\log g}}\right)^2\right) \\ &\cdot \exp\left(-\left(\frac{[\mathrm{Fe/H}]_\mathrm{s} - [\mathrm{Fe/H}]_0}{\sigma_{[\mathrm{Fe/H}]}}\right)^2\right), \end{aligned} \tag{2.33}$$

where $\{\theta_\mathrm{s}, \log g_\mathrm{s}, [\mathrm{Fe/H}]_\mathrm{s}\}$ are the parameters of the star and SN_s is the signal-to-noise ratio (SNR) of the star. The maximum SNR, SN_max, is defined to be the 99.7 percentile of the the SN ratios of all the stars. If $\mathrm{SN}_\mathrm{s} > \mathrm{SN}_\mathrm{max}$, SN_s is set equal to SN_max. Once we have a weight for each of the stars in a box, the spectrum of that box S_B is calculated as

$$S_\mathrm{B} = \frac{\sum_\mathrm{s} W_\mathrm{s} S_\mathrm{s}}{\sum_\mathrm{s} W_\mathrm{s}}, \tag{2.34}$$

with W_s the weight of star s and S_s its spectrum. In the same way, we also determine a corresponding set of of parameters p_B for each of the boxes, such that

$$p_\mathrm{B} = \frac{\sum_\mathrm{s} W_\mathrm{s} p_\mathrm{s}}{\sum_\mathrm{s} W_\mathrm{s}}, \tag{2.35}$$

with $p = \{\theta, \log g, [\mathrm{Fe/H}]\}$.

Each box that contain stars is assigned a weight W_j based on the distance of the box parameters p_B to the point p_0, so that

$$\begin{aligned} W_j = &\exp\left(-\left(\frac{\theta_\mathrm{B} - \theta_0}{\sigma_\theta}\right)^2\right) \cdot \exp\left(-\left(\frac{\log g_\mathrm{B} - \log g_0}{\sigma_{\log g}}\right)^2\right) \\ &\cdot \exp\left(-\left(\frac{[\mathrm{Fe/H}]_\mathrm{B} - [\mathrm{Fe/H}]_0}{\sigma_{[\mathrm{Fe/H}]}}\right)^2\right), \end{aligned} \tag{2.36}$$

and as a final step the spectrum S_0 of the point p_0 that we want to interpolate is calculated as

$$S_0 = \frac{\sum_j W_j S_{\mathrm{B},j}}{\sum_j W_j}, \tag{2.37}$$

where the sum runs over all of the boxes that contain stars. For more details we refer to Vazdekis et al. (2003).

2.3.3.1 Polynomial correction of the MILES stars

To test the interpolator, we have created an interpolated spectrum S_int for each of the stars in the MILES library, in such a way that the original MILES star was excluded from the data set that we use to build the interpolator. In this way, we calculate for each of the spectra in the MILES library the average residual R_S between the original spectrum S_or and the interpolated spectrum S_int. The residual is weighted by the average of the original spectrum, such that

$$R_\mathrm{S} = \left\langle \frac{\mathrm{abs}\left(S_\mathrm{or} - S_\mathrm{int}\right)}{S_\mathrm{or}} \right\rangle. \tag{2.38}$$

This allowed us to assess the quality of the interpolator and to identify stars with problems. A large mismatch between the interpolated spectrum

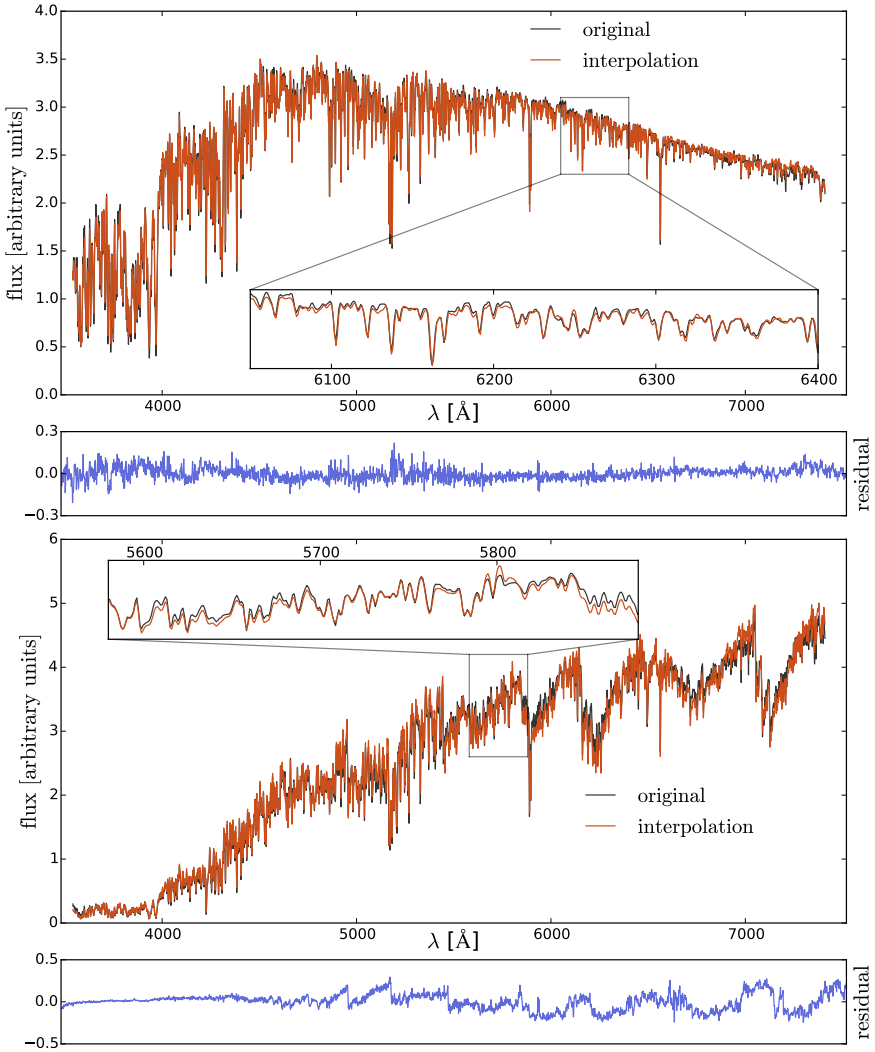

Figure 2.3 – Two examples where an observed MILES spectrum is compared to an interpolated spectrum with the same atmospheric parameters. The original MILES star is not part of the data set used to build the interpolator. The black lines represent the original spectra, the red lines the interpolated spectra. For both spectra, the residual between the original spectrum and the interpolated spectrum is shown in blue. **Top panel:** Interpolated spectrum MILES star with $T_{\rm eff} = 5392$ K, $\log g = 4.6$ and [Fe/H] = 0.1. **Bottom panel:** Interpolated spectrum MILES star with $T_{\rm eff} = 3793$ K, $\log g = 1.4$ and [Fe/H] = 0.32. The residual between the spectrum and the interpolated spectrum is on average 1.2% of the average of the original spectrum for the first spectrum and 2.9% for the second spectrum.

and the original spectrum may be the result of a low SNR of the original spectrum, any form of peculiarity in the original spectrum or using incorrect atmospheric parameters for the star in question. Problematic stars with a low signal-to-noise ratio, stars with obvious problems in their spectra after visual inspection and some peculiar stars that showed a large mismatch with their interpolated spectrum were removed from the dataset. Overall we removed 46 stars from the library.

A mismatch between the continuum of the original star and the interpolated spectrum may be caused by uncertainties in both the flux calibration and the correction for extinction. To absorb this effect, we correct each of the stars in the MILES library by a first order polynomial. This polynomial correction is an iterative process. At every step, the star with the largest residual is selected and corrected by a polynomial. Then the residuals are calculated again for the new data set. This process is repeated until each star has been corrected. Each star is corrected only once. After this correction, the average residual between the stars in the MILES library and the interpolated spectra is 2.3%. If we exclude stars with any notion of peculiarity in SIMBAD the average residual becomes 1.8%. Fig. 2.3 shows two examples of a comparison between a MILES spectrum and an interpolated spectrum with the same atmospheric parameters.

Having the interpolator in place, we create spectra for all the stars defined by the set of isochrones in the age-metallicity grid $\log \mathrm{age} = \{8.0 - 10.11\}$ yr and $[\mathrm{Fe/H}] = \{-1.0 - 0.4\}$. As a final step, the spectra resulting from the interpolator are scaled to match the V-magnitudes of the stars defined by the isochrone, for which we use the filter response defined in Maíz Apellániz (2006).

2.4 Results - mock single stellar populations

In this section we apply our model to a number of mock SSPs. We create the spectra for these mock SSPs by combining the stellar templates of one particular age and metallicity with an IMF.

To model the velocity dispersion of real stellar populations, we smooth the stellar templates to a velocity dispersion of 150 km s^{-1}. Before applying the model, we smooth the stellar templates to the same velocity dispersion. In Appendix 2.A we show that the results that we obtain for our mock SSPs do not depend on the velocity dispersion.

Note that although here we fix the velocity dispersion of the templates, this could in principle also be a free parameter of the model that we sample together with the other non-linear parameters. We will implement this in a future version of the code.

As a final step, we add Gaussian noise to the mock spectra to mimic an observation with a certain signal-to-noise ratio (SNR).

2.4.1 The regularization scheme

Before we apply the model to the mock SSPs we have to specify a regularization scheme $\mathbf{C}_{\rm pr}^{-1}$. For the mock SSPs we choose to use a regularization scheme that penalizes the relative deviation of the weights from the prior and prefers smooth deviations from the prior. This is expressed through the following inverse covariance matrix for the prior

$$\mathbf{C}_{\rm pr}^{-1} = \mathbf{C}_1 + \mathbf{C}_2, \tag{2.39}$$

where $\mathbf{C}_1$ is a diagonal matrix with

$$\mathbf{C}_{1_{i,i}} = \frac{1}{w_{0,i}^2}, \tag{2.40}$$

and $\mathbf{C}_2$ enforces smoothness of the deviations by using the following form of gradient regularization

$$\mathbf{C}_2 = \begin{pmatrix} \frac{1}{w_{0,1}^2} & -\frac{1}{w_{0,2}^2} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{w_{0,2}^2} & -\frac{1}{w_{0,3}^2} & 0 & \dots & 0 \\ 0 & 0 & \ddots & & & \vdots \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & \dots & \dots & \dots & \frac{1}{w_{0,n}^2} \end{pmatrix} \tag{2.41}$$

2.4.2 First level of inference

Within the first level of inference, we assume a certain model family and try to reconstruct its underlying parameters. Here we parametrize the IMF as a double power law with a break at 0.5 $M_\odot$. We split the reconstruction of the model parameters into a linear and a non-linear part. The linear part consists of the weights assigned to the individual stellar templates (i.e. the actual best-fitting IMF) whereas the non-linear part consists of determining

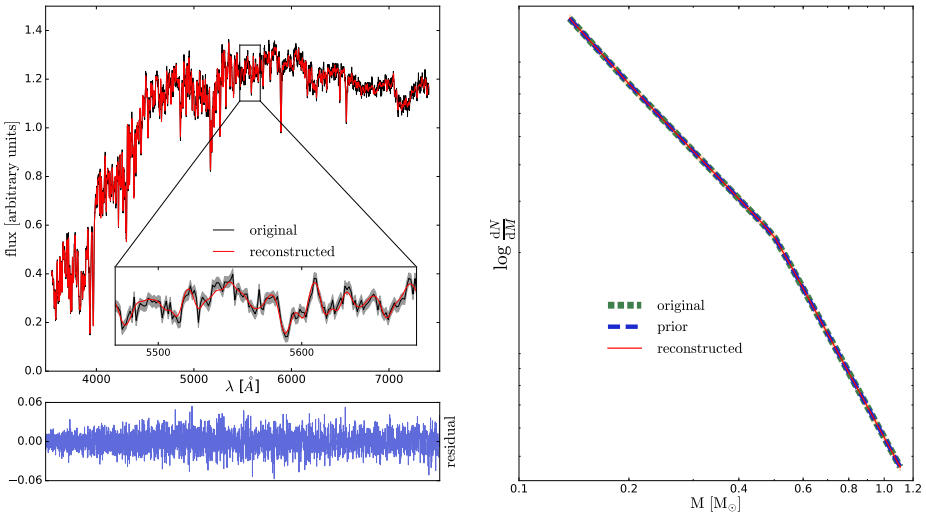

Figure 2.4 – Reconstructed spectrum and IMF of a mock SSP. The input spectrum is a mock SSP for which the underlying IMF is a double power law Kroupa IMF. We applied our model to the spectrum after adding random noise to this spectrum such that the SNR of the spectrum is 70. We provide the model with the correct set of stellar templates and specify the (correct) prior as a double power law Kroupa IMF. **Top panel:** Reconstructed spectrum. The *black* line shows the spectrum of the mock SSP and the *red* line corresponds to the spectrum reconstructed by the model. The residual between the two spectra is shown in *blue*. The *shaded gray* region in the zoom-in represents the one sigma uncertainty corresponding to the specified SNR. **Bottom panel:** Reconstruction of the IMF. The *short-dashed green* line represents the original IMF, the *long-dashed blue* line the prior IMF and the *red* line represents the reconstructed IMF by the model. The *shaded orange* region corresponds to the error on the weights resulting from the linear inversion, derived as described in Section 2.2.1.4 (almost invisible in this plot).

the most probable regularization parameter, finding the age and metallicity of the SSP and sampling the parameters of the IMF prior parametrization.

2.4.2.1 Linear parameters

The linear parameters in our model are represented by the weights $\boldsymbol{w}$ which correspond to the number of stars that we have for each of the templates in our model. As described in Section 2.2, these weights allow us to reconstruct the piecewise IMF of the stellar population.

To demonstrate the reconstruction of the IMF we consider a mock SSP with an age of 8.5 Gyr and solar metallicity. This mock SSP has an average

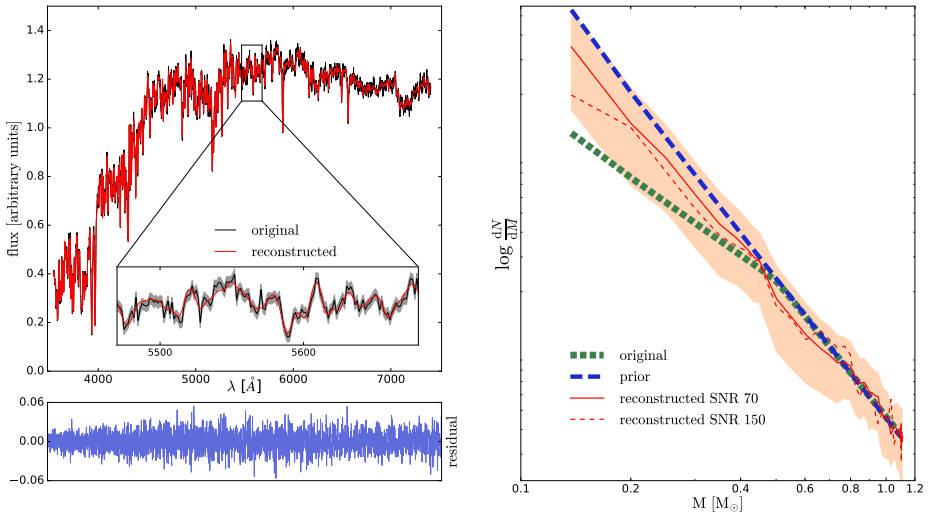

Figure 2.5 – As in Fig. 2.4, but when providing the model with an incorrect Salpeter IMF prior.

SNR of 70 and the underlying IMF of the stellar population is a double power law Kroupa IMF (i.e. a break at 0.5 $M_\odot$, a low-mass slope $\alpha_1 = 1.3$ and a high-mass slope $\alpha_2 = 2.3$).

To reconstruct the piecewise IMF, we first need to find the most probable distribution of weights $\boldsymbol{w}_{\rm MP}$ given a set of stellar templates and a prior IMF. If we fix the age and metallicity of the templates to the true values (which we know a priori in this case), the only thing that we can change in the model is the prior IMF.

First we consider what happens when we use the correct prior, a double power law Kroupa IMF (but let the level of regularization be free in the optimization). The upper panel of Fig. 2.4 shows the spectrum of the mock SSP together with the reconstructed spectrum of the model. The average of the spectrum divided by the standard deviation of the difference between the spectrum and the reconstructed spectrum is 68.6, consistent with an average SNR of 70. The lower panel of Fig. 2.4 shows the reconstructed IMF compared to the original IMF and the prior IMF. As one might expect, in this case the original IMF, the prior IMF and the reconstructed IMF all lie on top of each other.

As a next step, we provide the model with an (incorrect) Salpeter IMF prior (i.e. a constant slope of $\alpha_1 = \alpha_2 = 2.35$). The regularization

parameter is again free and optimized by the model. The reconstructed spectrum and reconstructed IMF that we obtain are shown in the upper and lower panel of Fig. 2.5. In this case the reconstructed IMF converges between the original IMF and the prior IMF, reflecting the fact that the minimization routine on the one hand wants to provide a good fit to the data and on the other hand wants to stay as close as possible to the prior.

Moreover, the obtained solution tends more and more towards the prior as we go to lower masses. This is probably related to the fact that the lowest mass stars contribute very little light to the spectrum. Therefore, it becomes much harder for the model to derive the abundance of these lower mass stars and hence the solution tends more and more towards the incorrect prior.

There also appears to be a degeneracy between the lower mass templates ($M \lesssim 0.4$ M$_\odot$) and the higher mass templates ($M \gtrsim 0.4$ M$_\odot$). This degeneracy is such that the model corrects the over-abundance of the lower mass templates with respect to the original weights (enforced by the prior) by decreasing the abundance of the templates that have slightly higher masses. In Fig. 2.6 we show the reconstructed spectrum from the most probable weights as compared to the original input spectrum without noise. This figure shows us that the obtained solution provides a very good fit to the data and it appears that this fit is as good as that obtained for the real solution within the uncertainty computed from the covariance matrix. Since this solution lies closer to the prior IMF, it is preferred over the actual solution. Except for the lowest mass bin, the actual solution lies within the one sigma errors of our solution.

One might expect that this degeneracy disappears if we increase the SNR. However, repeating this analysis for a number of (higher) SNRs shows that the degeneracy does not completely disappear. As an example, we show in the lower panel of Fig. 2.5 the reconstructed IMF for a spectrum with SNR=150. Although the obtained solution is now closer to the true one, apparently the data is not informative enough to break the degeneracy.

The final test that we discuss here is that of an IMF which is not part of the prior space allowed by the IMF parametrization. For this test we consider a 13 Gyr old mock SSP with solar metallicity. We assign weights to the different stellar templates based on a Kroupa IMF ($w_{\rm Kroupa}$) and then add to these weights a sinusoidal bump. This bump is added to the templates in the mass range 0.3 through 0.7 M$_\odot$ with a maximum of 10%

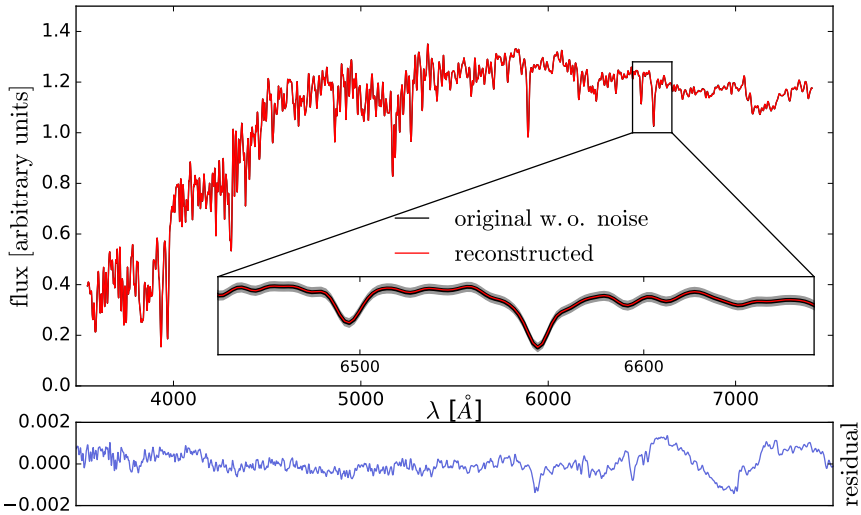

Figure 2.6 – Reconstructed spectrum of mock SSP from figure 2.5. We compare the reconstructed spectrum against the noise-free input spectrum. The figure shows us that the incorrect solution that we obtain in Figure 2.5 provides an excellent fit to the data. This illustrates that there exists a degeneracy between the stellar templates in the model.

for $M = 0.5\ \mathrm{M}_\odot$, so that we have

$$w(M) = w_{\mathrm{Kroupa}} \left(1 + 0.1 \cos\left(\frac{\pi}{0.4} M - 0.5\right)\right), \tag{2.42}$$

for templates in the range $M = [0.3, 0.7]\ \mathrm{M}_\odot$.

To reconstruct the IMF for this mock SSP we use a Kroupa IMF prior. The reconstructed spectrum and IMF are shown in Fig. 2.7. This figure shows that, although not by much, the obtained solution is different from the prior in the sense that it deviates slightly towards the true solution. And although this may not allow us to find the true solution, it provides a clear indication that the prior we used does not completely fit the data.

2.4.2.2 Non-linear parameters

The non-linear parameters of our model are represented by the age and metallicity of the stellar templates and by the parameters of the IMF prior model. We split the sampling of these parameters into two parts.

First we determine the evidence for a grid of ages and metallicities. Each point in this grid is associated with a set of stellar templates with

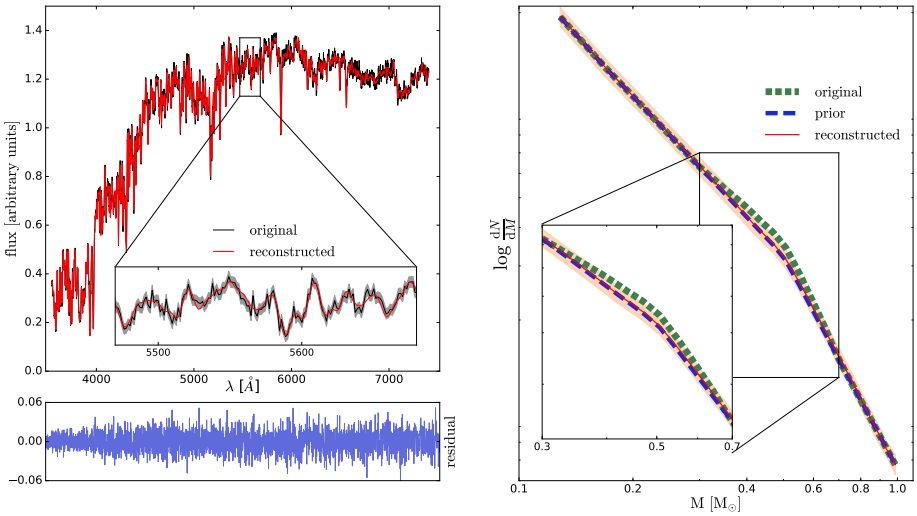

Figure 2.7 – As in Fig. 2.4, but now the underlying IMF has a bump between $M = 0.3$ $\mathrm{M}_\odot$ and $M = 0.7$ $\mathrm{M}_\odot$, as described in the text.

corresponding age and metallicity. We select the templates with the highest evidence. Secondly, we use these templates to refine the sampling of the parameters of the IMF prior model. Note that in principle `Multinest` already provides us with a sample of the parameters of the IMF prior and that the sampling of these parameters with `emcee` for the stellar templates with the highest evidence is only required to get a more precise sampling. Inside this loop, for every sample of the IMF prior model parameters p_i, the model continuously determines the most probable regularization parameter and solves for the most probable weights $\boldsymbol{w}_{\mathrm{MP}}$.

To demonstrate the reconstruction of the non-linear parameters we consider a set of twelve mock SSPs. The first nine mock SSPs have a Kroupa IMF ($\alpha_1 = 1.3$, $\alpha_2 = 2.3$) and the last three mock SSPs have a bottom-heavy IMF with $\alpha_1 = \alpha_2 = 3.0$ (the two low-mass slopes limit the range of low-mass slopes that are currently being considered in the literature as reasonable for early-type galaxies). All of the mock SSPs have a SNR of 70. The parameters of the different mock SSPs are summarized in Table 2.1.

Fig. 2.8 shows the results for the reconstruction of the non-linear parameters of `mock5`. The age-metallicity grid in this plot shows us that the stellar templates that we used as an input (i.e., $\log t[\mathrm{Gyr}] = 9.93$ and

Table 2.1 – Overview of the 12 mock SSPs considered in Section 2.4.2.2. The table provides both the input and reconstructed values of each mock SSP. In addition, we present the difference in log evidence with the second best set of stellar templates. In the column IMF, K represents a Kroupa IMF and BH a bottom-heavy IMF. The reconstructed IMF slopes are the median values of the distribution. The errors on the reconstructed slopes correspond to the distance between the median and the 16th and the 84th percentile. In the last two columns we show the maximum a posteriori (MAP) values of the sample for the two IMF slopes.

name	age	[Fe/H]	IMF	reconstructed		Δ evidence	reconstruction IMF prior parameters			
	[Gyr]			age [Gyr]	[Fe/H]	2nd best	$\alpha_{1,\mathrm{med}}$	$\alpha_{2,\mathrm{med}}$	$\alpha_{1,\mathrm{map}}$	$\alpha_{2,\mathrm{map}}$
`mock1`	3.1	-0.2	K	3.1	-0.2	15.8	$1.39^{+0.40}_{-0.73}$	$2.31^{+0.07}_{-0.07}$	1.50	2.32
`mock2`	3.1	0.0	K	3.1	0.0	16.0	$0.95^{+0.58}_{-1.40}$	$2.32^{+0.08}_{-0.08}$	1.41	2.34
`mock3`	3.1	0.2	K	3.1	0.2	7.5	$0.93^{+0.51}_{-0.97}$	$2.32^{+0.10}_{-0.08}$	1.20	2.27
`mock4`	8.5	-0.2	K	8.5	-0.2	9.7	$1.17^{+0.22}_{-0.40}$	$2.35^{+0.10}_{-0.10}$	1.34	2.32
`mock5`	8.5	0.0	K	8.5	0.0	8.0	$1.30^{+0.23}_{-0.30}$	$2.33^{+0.09}_{-0.09}$	1.29	2.34
`mock6`	8.5	0.2	K	8.5	0.2	11.9	$1.33^{+0.21}_{-0.27}$	$2.24^{+0.09}_{-0.09}$	1.37	2.23
`mock7`	13.0	-0.2	K	13.0	-0.2	7.8	$1.33^{+0.16}_{-0.39}$	$2.43^{+0.12}_{-0.10}$	1.40	2.41
`mock8`	13.0	0.0	K	13.0	0.0	10.0	$1.35^{+0.24}_{-0.29}$	$2.14^{+0.12}_{-0.10}$	1.41	2.12
`mock9`	13.0	0.2	K	13.0	0.2	4.6	$1.46^{+0.21}_{-0.21}$	$2.20^{+0.10}_{-0.10}$	1.38	2.22
`mock10`	3.1	0.2	BH	3.1	0.2	3.9	$2.95^{+0.07}_{-0.10}$	$3.00^{+0.07}_{-0.07}$	2.99	2.97
`mock11`	8.5	0.0	BH	8.5	0.0	9.1	$3.03^{+0.08}_{-0.09}$	$2.91^{+0.08}_{-0.08}$	3.06	2.90
`mock12`	13.0	-0.2	BH	13.0	-0.2	5.5	$3.00^{+0.13}_{-0.07}$	$3.03^{+0.09}_{-0.12}$	2.98	3.06

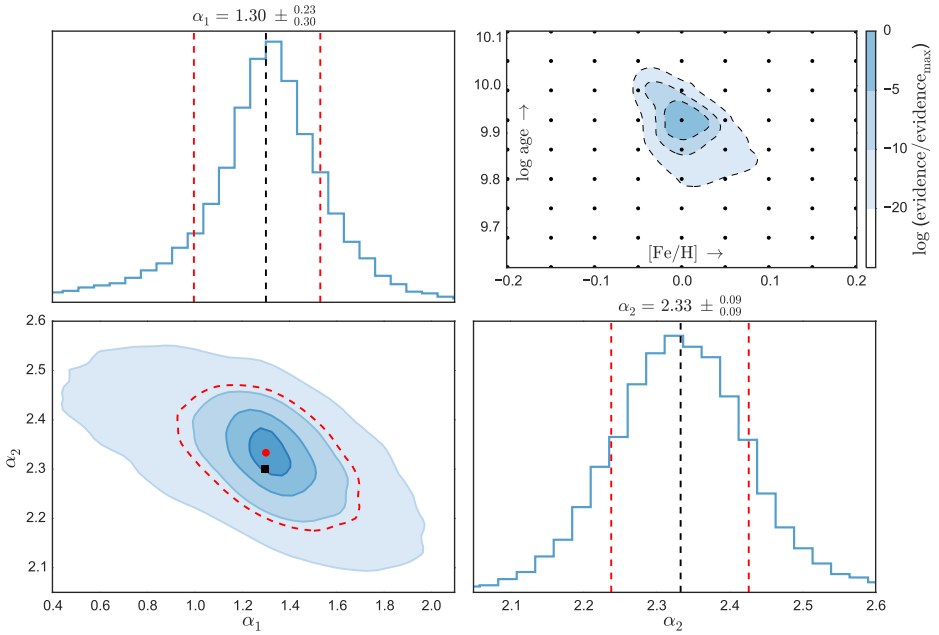

Figure 2.8 – Reconstructed parameters for `mock5`. **Upper-right panel:** log evidence obtained for the different stellar templates in the age-metallicity grid. The evidences are re-scaled such that the log evidence of the templates with the highest evidence is zero. **Lower-left panel:** two-dimensional plot of the probability density distribution resulting from sampling the IMF slopes α_1 and α_2. For the sampling we used the stellar templates with the highest evidence. The different colors contain 10%, 33%, 60% and 90% of the points in the sample. The red-dashed line corresponds to the 1σ confidence interval. The red dot corresponds to the median of the sample whereas the black square corresponds to the input Kroupa IMF. **Upper-left panel:** marginalized distribution for the low-mass slope α_1. **Lower-right panel:** marginalized distribution for the high-mass slope α_2. The black-dashed lines in the histograms correspond to the median values and the red-dashed lines to the 16th and 84th percentiles of the marginalized distribution.

[Fe/H] = 0.0) give us the highest evidence. Although the grid shows an age-metallicity degeneracy, the evidence difference with the other grid points is at least more than 8. This implies that there is substantial evidence in favour of the correct stellar templates and that we are able to reconstruct the age and metallicity of the mock SSP. Note that in the future we plan to further expand the sampling to cover the full space of age, metallicity and IMF prior parameters.

In addition to the age-metallicity grid, Fig. 2.8 also shows the reconstruction of the two IMF slopes. These slopes are part of the parameters

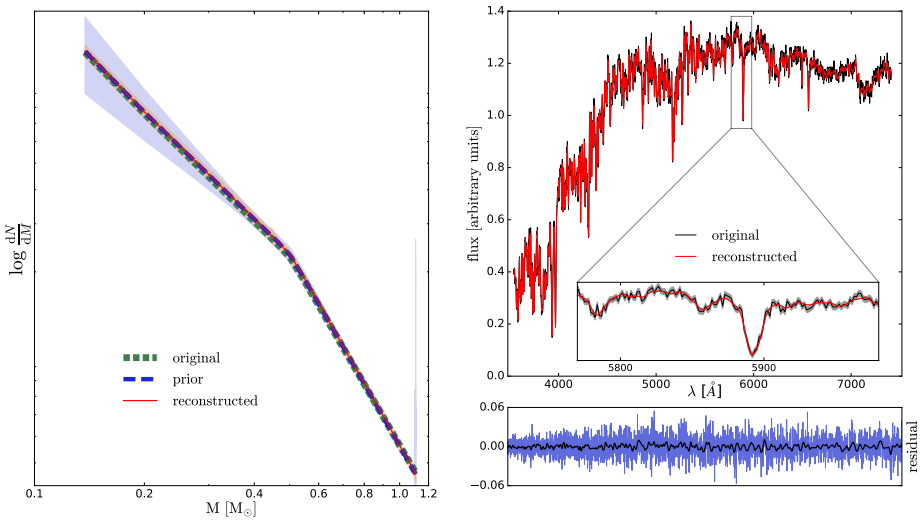

Figure 2.9 – Reconstruction of the piecewise IMF and the spectrum for `mock5`. **Top panel:** Reconstructed piecewise IMF for `mock5`. The *short-dashed green* line represents the original IMF, the *long-dashed blue* line represents the prior IMF, and the *red* line represents the IMF reconstructed by the model. The *shaded orange* region corresponds to the error on the weights from the linear inversion. The *shaded blue* region represents the area between the 16th and 84th percentile in the distribution of most probable weights. **Middle panel:** The *black* line corresponds to the spectrum of the mock SSP and the *red* line corresponds to the spectrum reconstructed by the model. The *shaded gray* region in the zoom-in represents the one-sigma uncertainty corresponding to the SNR of the spectrum. **Bottom panel:**, the *blue* line represent the residual between the original spectrum and the reconstructed spectrum. In the same panel, the *black* line represents the residual smoothed with a Gaussian kernel with $\sigma = 5$ pixels.

that define the assumed double power law IMF prior parametrization. The values presented for α_1 and α_2 in Fig. 2.8 correspond to the median values of the marginalized distributions whereas the errors represent the difference between the median and the 16th and 84th percentile. Within these confidence limits, the reconstructed IMF prior parameters for `mock5` agree well with the input parameters.

Once we have a set of best fit values[3] for the parameters of the IMF prior model, we use these parameters to construct a prior for the IMF. This prior may then be used to reconstruct the piecewise IMF, similar to what we did in Section 2.4.2.1. The reconstructed IMF for `mock5` is shown in

[3]We use the maximum a posteriori values of the sampled IMF prior parameters.

Fig. 2.9. Also shown in this figure is the reconstructed spectrum obtained by using this IMF together with the most probable stellar templates.

The plots corresponding to the other mock SSPs in Table 2.1 are shown in Appendix 2.B. For all of the mock SSPs, the reconstructed parameters are summarized in Table 2.1. We are able to select the true age and metallicity for all of our mock SSPs. As a measure of the robustness of this reconstruction, we present in Table 2.1 the difference in log evidence with the second best set of stellar templates. These numbers show us that for `mock3`, `mock4`, `mock5`, `mock7`, `mock9`, `mock10`, `mock11` and `mock12` there is substantial evidence in favour of the true set of stellar templates. For `mock1`, `mock2`, `mock6` and `mock8` there is strong evidence in favour of the true stellar templates. See also Section 2.2.1.6 for the interpretation of the difference in log evidence.

The reconstructed IMF slopes for the twelve mock SSPs are listed in Table 2.1. Except for `mock7`, `mock8` and `mock11`, we are able to reconstruct the input slopes of the IMF within the confidence limits resulting from the sampling procedure. For `mock7`, `mock8` and `mock11`, the true high-mass slope α_2 is just outside the confidence limits. Since the confidence limits correspond to the 16th and 84th percentile of the distribution, one would expect that indeed about one third of the test cases will be outside of these limits.

For the intermediate and older populations with a Kroupa IMF, the errors on the low-mass slope α_1 are significantly smaller than those of the younger populations with a Kroupa IMF. Although the absolute signal of the low-mass stars in an old and a young population may be the same, the additional light of the young stars that are still present in the younger population effectively reduces the SNR of the low-mass stars. So the younger an SSP, the lower the SNR of the low-mass stars (for a spectrum with a given SNR) and the more difficult it becomes to constrain the low-mass slope α_1.

Driven by the increasing error on α_1 for the mock SSPs with a Kroupa IMF, we also consider the fractional contribution $L_{0.5}$ of low-mass stars (i.e. $M < 0.5$ $\mathrm{M}_\odot$) to the integrated spectrum across the MILES wavelength range. Table 2.2 provides an overview of this fraction for the different ages and IMF prior models that we consider in Table 2.1 (for solar metallicity). We conclude that the signal from the low-mass stars in the youngest populations with a Kroupa IMF becomes comparable to or even lower than the intrinsic noise of the spectrum. Hence, this explains why there is a

Table 2.2 – Fractional contribution $L_{0.5}$ of low-mass stars ($M < 0.5$ $M_\odot$) to the integrated spectrum across the MILES wavelength range. These fractions are derived for the ages and IMF prior models considered in Table 2.1 for solar metallicity using our models.

age [Gyr]	IMF	$L_{0.5}$
3.1	Kroupa	0.9%
8.5	Kroupa	2.2%
13.0	Kroupa	3.2%
3.1	bottom-heavy	4.2%
8.5	bottom-heavy	8.6%
13.0	bottom-heavy	11.3%

sudden increase in the error on α_1 if we go from the intermediate to the youngest populations and why the effect is much smaller if we go from the oldest populations to the intermediate populations.

By increasing the age of an SSP, one reduces light from the more massive stars in the SSP and effectively increases the SNR of the low-mass stars. A different way to increase the SNR of the low-mass stars is to increase the number of low-mass stars. To realize this, we consider a set of three bottom-heavy SSPs (`mock10`, `mock11` and `mock12`). Table 2.2 shows that the relative contribution of low-mass stars to the integrated spectrum is much higher than it is for a Kroupa IMF. This is reflected in the smaller errors on the low-mass slope α_1 in Table 2.1, even for the youngest population where the signal of the low-mass stars is still well above the intrinsic noise of the spectrum.

The high-mass slope α_2 is, independently of age, determined by the stars that emit most of the light. Therefore we expect that the model accurately reconstructs α_2 for all mock SSPs. This is confirmed by the relatively small errors on α_2 in Table 2.1, which are more or less constant as a function of age.

All of the two-dimensional probability density plots in Fig. 2.8 and Appendix 2.B show a clear degeneracy between the low-mass slope α_1 and high-mass slope α_2. This degeneracy is such that an increasing low-mass slope seems to prefer a decreasing high-mass slope. To interpret this result we show in Fig. 2.10 two reconstructed IMFs corresponding to the one sigma values around the medians of the sampled values. For the first reconstruction we combine a lower low-mass slope with a higher high-

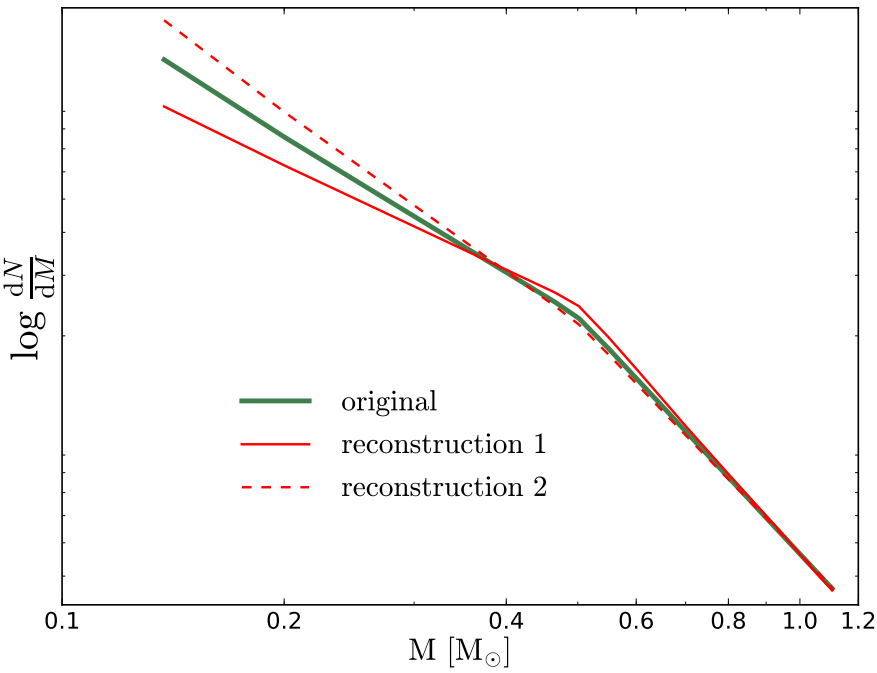

Figure 2.10 – Two different IMF reconstructions for `mock5`. The first reconstruction combines a lower low-mass slope with a higher high-mass slope and the second reconstruction combines a higher low-mass slope with a lower high-mass slope. These reconstructions correspond to the one sigma values around the median of the sampled IMF slopes and reflect the degeneracy between the IMF slopes in Fig. 2.8.

mass slope (i.e. $\alpha_1 = 1.0$, $\alpha_2 = 2.42$) and for the second reconstruction we combine a higher low-mass slope with a lower high-mass slope (i.e. $\alpha_1 = 1.53$, $\alpha_2 = 2.24$). The parameters of these two reconstructions reflect the degeneracy visible in Fig. 2.8. The normalization of the IMF is lower for the second reconstruction. For both reconstructions the model appears to correct an over-abundance (under-abundance) of the lowest mass templates with respect to the real IMF by an under-abundance (over-abundance) of the templates around 0.5 $M_\odot$ (luminosity conservation). In fact, the observed degeneracy between the IMF slopes may therefore be closely related to the degeneracy observed in Fig. 2.5.

2.4.2.3 Mass fraction of low-mass stars

Stellar templates in the same stellar mass range can have spectra that look very similar. As a consequence, there may be degeneracies between stellar templates with similar masses. Our regularization scheme ensures that these degeneracies do not cause problems when we solve for the most probable weights. However, if such a degeneracy exists it is basically impossible to find the exact contribution of the degenerate templates and hence the balance between these templates is set by the parametrization of the IMF prior. Although in that case we may not be able to completely reconstruct the piecewise IMF, we may still be able to constrain broader IMF-related quantities, for example the dwarf-to-giant ratio.

One of the important questions that we try to answer with our model is what the relative importance of low-mass stars is to the total stellar mass. In that context, transforming the most probable weights determined by our model into a fractional contribution of low-mass stars to the total stellar mass allows for a simple interpretation of the results.

La Barbera et al. (2013) define the fraction of the total initial stellar mass in stars with $M < 0.5\,\mathrm{M}_\odot$ as

$$\mathrm{Fraction}(< 0.5\mathrm{M}_\odot) \equiv \frac{\int_{0.1\,\mathrm{M}_\odot}^{0.5\,\mathrm{M}_\odot} \xi(M) M \mathrm{d}M}{\int_{0.1\,\mathrm{M}_\odot}^{100\,\mathrm{M}_\odot} \xi(M) M \mathrm{d}M}. \tag{2.43}$$

However, the SSPs that we consider here do not have stars more massive than $\sim 1.5\ \mathrm{M}_\odot$. Everything beyond the high-mass-cut-off (HMCO) of the current mass function (i.e. the highest mass star in an SSP of a given age and metallicity that is still present) is more sensitive to the parametrization of the IMF than to the actual distribution of stellar masses. This is particularly true if we extrapolate our reconstructed IMFs, which can possibly be irregular. Therefore, we define the quantity $F_{0.5}$ as the fraction of the total current stellar mass that is in stars with $M < 0.5\ \mathrm{M}_\odot$

$$F_{0.5}(< 0.5\mathrm{M}_\odot) \equiv \frac{\int_{0.1\,\mathrm{M}_\odot}^{0.5\,\mathrm{M}_\odot} M \xi(M) \mathrm{d}M}{\int_{0.1\,\mathrm{M}_\odot}^{m_{\mathrm{HMCO}}} M \xi(M) \mathrm{d}M}. \tag{2.44}$$

In Table 2.3 we summarize the mass fractions $F_{0.5}$ for our twelve mock SSPs. As a reference, for every SSP we report $F_{0.5,\mathrm{original}}$: the value of $F_{0.5}$ for an SSP with the same age and metallicity and an IMF equal to the input IMF. The results in Table 2.3 show that for all of our mock SSPs

Table 2.3 – Mass fraction of stars with $M < 0.5$ $M_\odot$ for the twelve mock SSPs in Table 2.1. The mass fractions and corresponding errors are derived from the most probable distribution of weights. As a reference, for every mock SSP we also provide $F_{0.5,\mathrm{original}}$, the value of $F_{0.5}$ that the corresponding SSP would have for a the input IMF (i.e. Kroupa or bottom-heavy).

name	$F_{0.5,\mathrm{original}}$	$F_{0.5}$
`mock1`	0.52	$0.55^{+0.05}_{-0.13}$
`mock2`	0.51	$0.53^{+0.02}_{-0.23}$
`mock3`	0.51	$0.49^{+0.05}_{-0.15}$
`mock4`	0.58	$0.59^{+0.01}_{-0.08}$
`mock5`	0.58	$0.58^{+0.03}_{-0.04}$
`mock6`	0.58	$0.58^{+0.03}_{-0.04}$
`mock7`	0.62	$0.64^{+0.02}_{-0.06}$
`mock8`	0.61	$0.61^{+0.02}_{-0.05}$
`mock9`	0.61	$0.61^{+0.04}_{-0.02}$
`mock10`	0.81	$0.80^{+0.01}_{-0.01}$
`mock10`	0.83	$0.83^{+0.01}_{-0.01}$
`mock12`	0.85	$0.85^{+0.01}_{-0.01}$

the value of $F_{0.5,\mathrm{original}}$ lies within one sigma of the reconstructed $F_{0.5}$ and these results are therefore consistent with the input data.

2.4.2.4 The signal-to-noise ratio

The spread that we find in the reconstructed parameters for the mock SSPs is related to the SNR of the input spectra. To demonstrate this, we compare the reconstructed parameters for four mock SSPs with different SNRs. We will consider a young, metal-rich population with the same parameters as `mock3` and an old, metal-poor population with the same parameters as `mock7`. For both populations we have a spectrum with a SNR=70 and a spectrum with a SNR=150. The spectra are summarized in Table 2.4.

The reconstructed IMF prior parameters (α_1, α_2 and the normalization of the IMF) for the mock SSPs in Table 2.4 are shown in Fig. 2.11. This figure shows that increasing the SNR from 70 to 150 decreases the

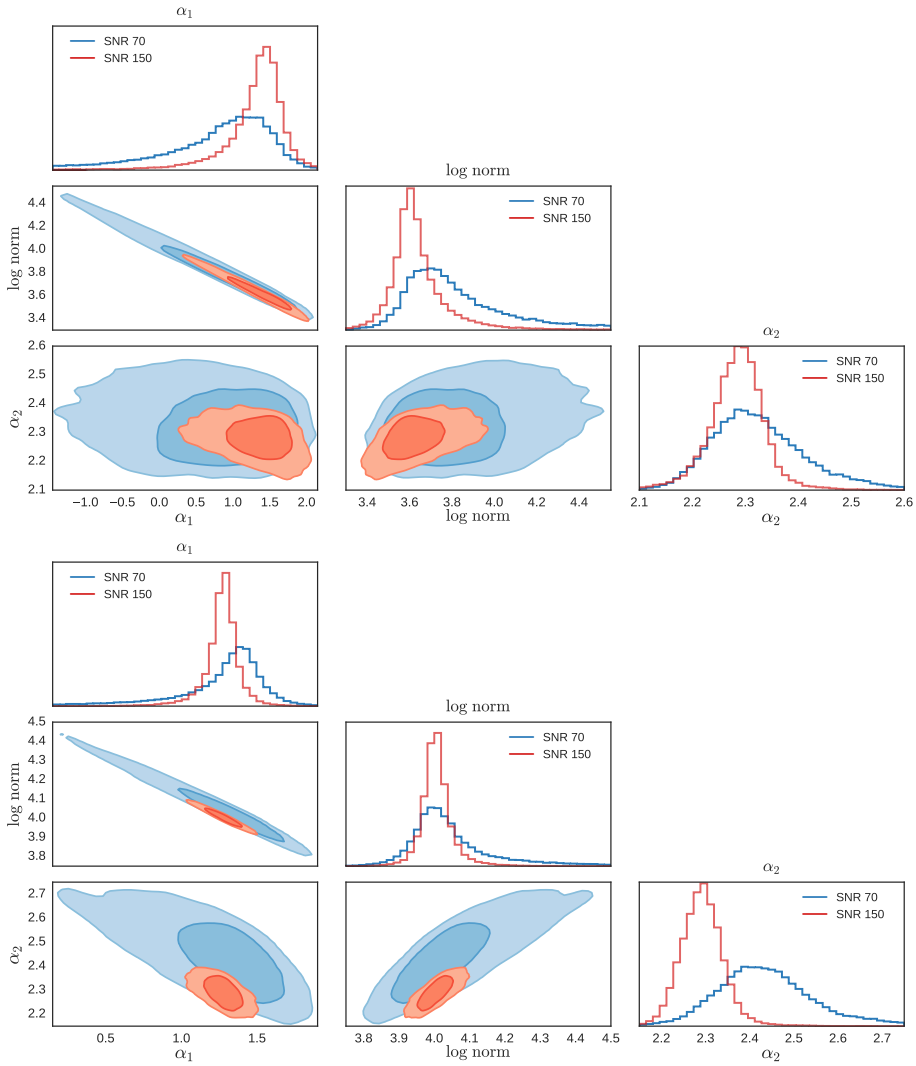

Figure 2.11 – The effect of the SNR on the reconstructed IMF parameters for the mock SSPs in Table 2.4. In the two-dimensional probability density plots, the *dark* and *light blue* contours contain 68% and 90% of the sampled points for the spectra with SNR=70. The *dark* and *light red* contours correspond to 68% and 90% of the sampled points for the spectra with SNR=150. Also shown are the marginalized distributions for the IMF parameters: *blue* corresponds to a SNR=70 and *red* to a SNR=150. **Top panel:** Reconstructed IMF parameters for `mock3-70` and `mock3-150`. **Bottom panel:** Reconstructed IMF parameters for `mock7-70` and `mock7-150`.

Table 2.4 – The effect of the SNR on the reconstructed parameters. For two of the mock SSPs in Table 2.1 we compare the reconstructed parameters for spectra with a SNR of 70 and spectra with a SNR of 150.

name	age [Gyr]	[Fe/H]	SNR	α_1	α_2
`mock3-70`	3.1	0.2	70	$0.93^{+0.51}_{-0.97}$	$2.32^{+0.10}_{-0.08}$
`mock3-150`	3.1	0.2	150	$1.38^{+0.23}_{-0.39}$	$2.29^{+0.04}_{-0.05}$
`mock7-70`	13.0	-0.2	70	$1.33^{+0.16}_{-0.39}$	$2.43^{+0.12}_{-0.10}$
`mock7-150`	13.0	-0.2	150	$1.28^{+0.08}_{-0.09}$	$2.29^{+0.04}_{-0.05}$

spread in the reconstructed IMF prior parameters and results in a much sharper peak in the marginalized distributions. In this case, the effect of increasing the SNR is very clear. However, for these mock SSPs the only source of uncertainty that we have is the Gaussian noise that we add a priori to the spectra of the SSPs. When we consider real data, there are additional problems and uncertainties that play a role, including flux-calibration issues, telluric residuals, abundance ratios and systematic uncertainties between the interpolated stellar templates and the true stellar templates. These errors are non-negligible and therefore increasing the SNR will not always improve the reconstruction of the IMF model parameters.

2.4.2.5 Wavelength dependence

The residuals between the original spectrum and the reconstructed spectrum for `mock5` in Fig. 2.9 do not show a strong dependence on wavelength. To investigate a possible wavelength dependence of the residuals in more detail, we smooth the residuals with a Gaussian kernel with $\sigma = 5$ pixels. The smoothed version of the residual is also shown in Fig. 2.9. The smoothed residuals seem to be slightly smaller at the bluest wavelengths, but this is mainly related to the lower flux in that region (the residuals are absolute).

Although we do not see a wavelength dependence in the residuals of our results, this does not tell us which wavelength regions are better suited to constrain the IMF. To investigate the relation between the reconstruction of the IMF prior parameters and the considered wavelength region we repeat our analysis for `mock5` using only the blue half of the spectrum and using only the red half of the spectrum. The results of this analysis are presented in Table 2.5.

Table 2.5 – Effect of wavelength region on the reconstruction of the IMF slopes α_1 and α_2 for `mock5`.

wavelength region	α_1	α_2
3540-7410 Å	$1.30^{+0.23}_{-0.30}$	$2.33^{+0.09}_{-0.09}$
3540-5122 Å	$-0.59^{+1.15}_{-0.96}$	$2.54^{+0.14}_{-0.16}$
5122-7410 Å	$0.99^{+0.53}_{-0.98}$	$2.23^{+0.17}_{-0.18}$

Firstly, the results in Table 2.5 show that ignoring half of the data points significantly increases the scatter in the reconstructed parameters. Secondly, the reconstructed parameters that we obtain using the reddest wavelength region are consistent with the input data whereas the reconstructed parameters derived from the bluest wavelength region are not. Since the overall flux in the blue part of the spectrum is lower, the SNR of the blue part is also slightly lower. Moreover, according to our models in the blue part of the spectrum the low-mass stars contribute only 1.5% to the integrated spectrum whereas in the red part they contribute 2.6%. Taking into account that these numbers have the same order of magnitude as the intrinsic noise in the spectrum this represents an important difference. This difference and the lower SNR in the blue part of the spectrum might explain why we are able to constrain α_1 correctly in the red part of the spectrum and not in the blue part of the spectrum. The offset that we find for α_2 in the blue part of the spectrum may be related to the model being unable to correctly break the degeneracy between the IMF slopes.

The results in this section show that in order to constrain the contribution of low-mass stars to the spectrum it is important to consider a broad wavelength region. Furthermore, it is essential to consider wavelength regions and features that are sensitive to changes in the IMF. Since low-mass stars emit most of their light in the (infra)red part of the spectrum and most of the IMF sensitive features are also found in this part of the spectrum (Spiniello et al. 2014), (infra)red wavelength regions will be more useful to constrain the low-mass IMF. In future work, we therefore plan to combine our model with the X-Shooter Spectral Library which provides a much broader and redder wavelength coverage.

Table 2.6 – Evidence for two different model families obtained by applying it to two different mock SSPs. The mock SSPs are created with the same stellar templates. One of the mock SSPs has a Kroupa IMF and the other has a Salpeter IMF. We compare a single power law IMF model family against a double power law IMF model family.

true IMF	IMF model family	log evidence
Kroupa	single power law	-22085.4 ± 0.2
Kroupa	double power law	-22083.0 ± 0.2
Salpeter	single power law	-22049.3 ± 0.2
Salpeter	double power law	-22051.7 ± 0.2

2.4.3 Second level of inference

The second level of inference allows us to compare different model families based on the evidence. Here we consider two model families: one in which the IMF is parametrized as a double power law and one in which it is parametrized as a single power law. For simplicity, although there could in principle also be a degeneracy between age/metallicity and IMF model family, for now we fix the age and metallicity in both model families to the true values, such that the stellar templates are the same for the two model families.

To test the second level of inference we consider two mock SSPs with $t = 8.5$ Gyr and [Fe/H] $= 0.0$. The SNR of the spectra of these SSPs is 70. One of the SSPs has a Kroupa IMF and the other a Salpeter IMF.

We then use `Multinest` to calculate the evidence for both model families by applying them to the two mock SSPs. The results are presented in Table 2.6.

First consider the Kroupa mock SSP. The difference in log evidence is 2.4 in favour of the double power law model family. Since a broken power law IMF is not part of the model family of single power laws, this is what we expect. However, the difference in log evidence between the two model families is not substantial. Apparently there is a single power law that, in combination with the allowed deviations from the prior IMF, is still able to provide a reasonable fit to the data. If this would not have been the case, the difference in the evidence would have been much more significant.

For the Salpeter mock SSP, the difference in log evidence is 2.4 in favour of of the single power law model family. In this case the single power law input IMF is part of both model families. Therefore we expect the two

model families to be able to fit the data equally well. Nevertheless, the model prefers the simpler single power law model. This is the result of Occam's razor: the model is set up in such a way that it tries to find the simplest model that fits the data.

Considering the errors on the evidences in Table 2.6, we conclude that for these specific mock SSPs we are able to discriminate between the single- and double-power-law model families. The difference in log evidence between the two model families is however not substantial.

2.5 Results - model versus model

As a next step, we apply our model to a set of SSPs that have been constructed with a different SPS code. Because the current version of our model is based on the MILES library, the obvious choice is to compare our model against the MILES SPS models (Vazdekis et al. 2010, 2015). Here we consider three SSPs. The parameters of these SSPs are given in Table 2.7. We downloaded the spectra of these SSPs from the MILES website

Before we discuss the reconstruction of the IMF for these three SSPs we first specify the regularization scheme that we used for the analysis of the MILES spectra.

2.5.1 The regularization scheme

For the MILES mock SSPs, there are also systematic uncertainties between the model and the input spectra in addition to the Gaussian noise that we add to the spectra. Since we do not take these uncertainties into account in the covariance matrix, sometimes this may result in unrealistic IMFs. This is a consequence of the model trying to provide a fit to the data that is too good with respect to the real uncertainties and at the cost of large deviations from the prior IMF (i.e. if the regularization parameter is low; this is also related to the use of NNLS). We solve this problem for now by using a different regularization scheme. For the MILES SSPs, we use $\mathbf{C}_{\rm pr}^{-1} = \mathbf{I}$ (i.e. the identity matrix). This regularization scheme penalizes the absolute deviation of the weights from the prior whereas the regularization scheme that we used in Section 2.4 penalizes both the relative deviation of the weights from the prior and the gradient of the relative deviations. In general there are much more low-mass stars than high-mass stars. Therefore the regularization scheme that we use in this section makes it harder for

Table 2.7 – Overview of the three MILES SSPs considered in Section 2.5. All of these mock SSPs have been downloaded from the MILES website and have a SNR of 70. The table provides both the input parameters and reconstructed values for each of the mock SSPs. In addition, we present the difference in log evidence with the second best set of stellar templates. The reconstructed IMF slopes are the median values of the distribution. Errors on the reconstructed slopes correspond to the 16th and the 84th percentile. The last two rows represent the reconstructed IMF slopes using a set of stellar templates from the MILES website.

name	`MILES1`	`MILES2`	`MILES3`
parameters			
age [Gyr]	3.2	8.9	12.6
[Fe/H]	0.22	0.0	-0.40
IMF	Kroupa	Kroupa	Kroupa
reconstructed parameters			
age [Gyr]	3.6	9.8	13.0
[Fe/H]	0.30	-0.05	-0.40
Δ evidence 2nd best	4.0	4.2	4.3
α_1	$3.87^{+0.36}_{-0.44}$	$2.08^{+0.44}_{-0.61}$	$1.57^{+0.58}_{-0.71}$
α_2	$0.89^{+0.27}_{-0.30}$	$2.16^{+0.20}_{-0.20}$	$1.97^{+0.30}_{-0.29}$
$\alpha_{1,\mathrm{MILES}}$	$-0.23^{+1.23}_{-1.17}$	$1.01^{+0.54}_{-0.69}$	$1.42^{+0.43}_{-0.52}$
$\alpha_{2,\mathrm{MILES}}$	$2.44^{+0.14}_{-0.14}$	$2.34^{+0.12}_{-0.15}$	$2.21^{+0.22}_{-0.24}$

the model to deviate from the prior for the low-mass templates and helps to prevent strange solutions that are not regulated strongly enough. However, it comes at the cost of a less flexible model for the low-mass end.

2.5.2 IMF reconstruction of MILES SSPs

We analyse the spectra of the three MILES mock SSPs in the same way as in Section 2.4, except for the different regularization scheme. We first calculate the evidence for every set of stellar templates in the age-metallicity grid. Then we refine the sampling of the non-linear IMF prior parameters for the stellar templates with the highest evidence.

The top panel of Fig. 2.12 shows the reconstructed parameters for the SSP with $t = 8.9$ Gyr and [Fe/H] $= 0.0$ (`MILES2`). The preferred set of stellar templates has an age of 9.8 Gyr and a metallicity [Fe/H] $= -0.05$.

Note that the age-metallicity grid in our model does not contain templates with $t = 8.9$ Gyr. The stellar templates in our age-metallicity grid that are closest in age have $t = 8.5$ Gyr and $t = 9.8$ Gyr. Although the selected stellar templates are slightly too metal-poor, the reconstructed age and metallicity agree reasonably well with the input values.

The reconstructed high-mass slope for `MILES2` is consistent with the input values of a Kroupa IMF. The value obtained for the low-mass slope is, on the other hand, too high and just outside the one sigma confidence limits given by the model. Notice, however, we only take into account the noise in the data in the covariance matrix $\mathbf{C}_{\mathrm{D}}$ and that systematic uncertainties are not taken into account. Therefore we expect the errors returned by the model to be a lower limit.

The bottom panel of Fig. 2.12 shows the reconstructed IMF for `MILES2`. The reconstructed IMF and the input IMF are consistent at the high-mass end. However, the original IMF for the low-mass end is just outside the one sigma confidence limits of the model. Note that the spike in the reconstructed IMF at the high-mass end corresponds to a template with a very small mass bin and that this spike is not visible in the reconstructed distribution of weights. For more details see also Appendix 2.C.

For the other two MILES SSPs, the results are shown in Appendix 2.C. The reconstructed parameters are summarized in Table 2.7. For all MILES SSPs, the reconstructed age and metallicity are relatively close to the input values.

Although the high-mass slope is slightly too low, the reconstructed IMF slopes for `MILES3` are consistent with a Kroupa IMF. The reconstructed slopes for `MILES1`, on the other hand, are inconsistent with the input values. There are a number of differences between the stellar templates in our model and the stellar templates in the MILES models that might explain the inconsistent reconstructions. First of all, we use a different set of isochrones. Second, although the interpolation mechanisms in the two models are similar they are not the same. In the MILES models there is an additional correction term that minimizes the difference between the requested stellar parameters and the stellar parameters of the interpolated spectrum whereas in our model there is a polynomial correction as described in Section 2.3.3.1. Finally, the MILES models use empirical color-temperature relations to scale the stellar templates, whereas we use the colors provided by the isochrones.

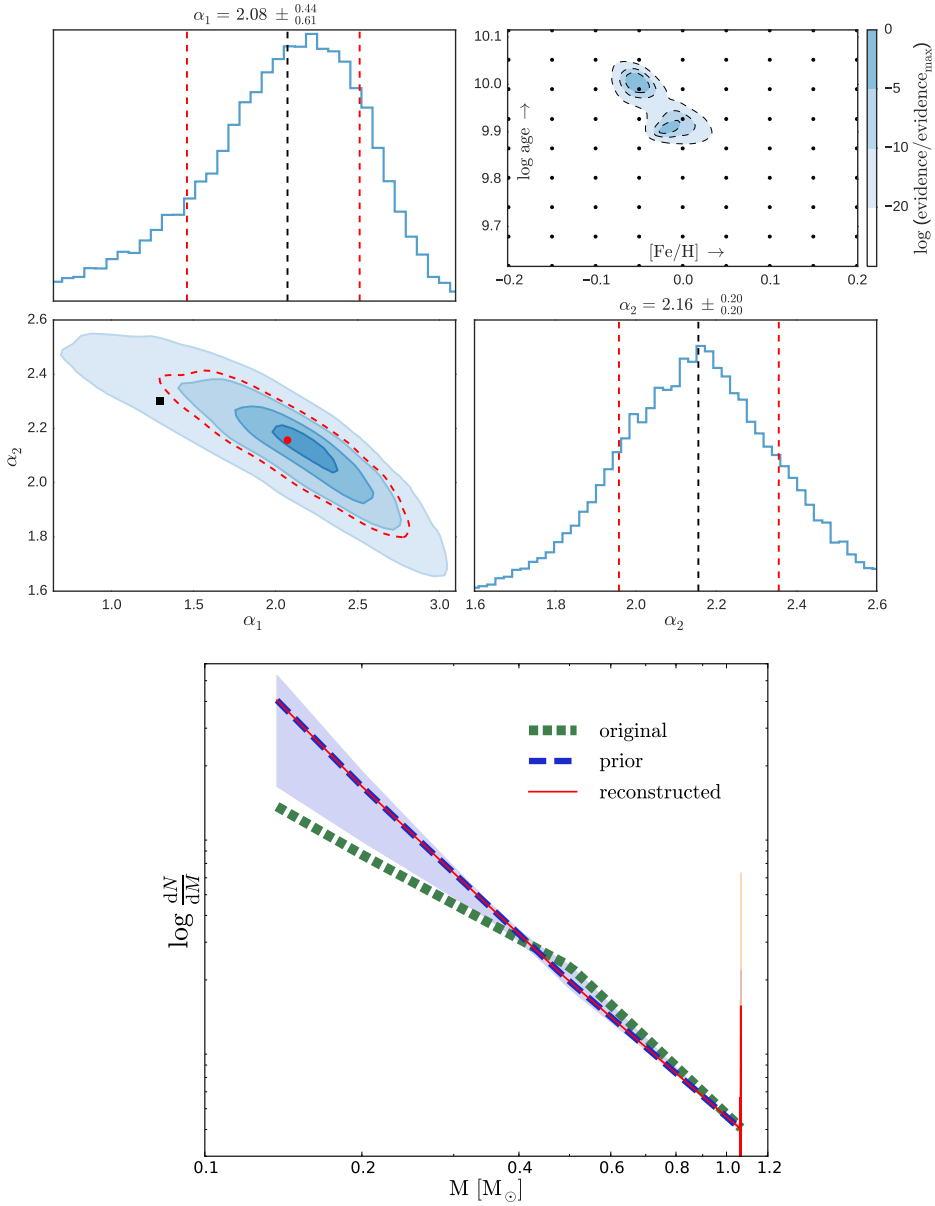

Figure 2.12 – Sampling of the IMF prior model parameters and reconstruction of the IMF for `MILES2`. **Top panel:** As in Fig. 2.8 only now for `MILES2`. **Bottom panel:** As in the top panel of Fig. 2.9 only now for `MILES2`.

We suspect that the stellar interpolator is one of the most important sources of uncertainty responsible for the discrepancy between the input IMF slopes for the MILES SSPs and the slopes reconstructed by our model. To test this hypothesis, we select an isochrone with $t = 8.9$ Gyr and [Fe/H] = 0.0. For all of the stars in this isochrone, we create a spectrum with the MILES interpolator by using the webtool on the MILES website. Using this set of stellar templates we once again apply our model to `MILES2`. The results are shown in Fig. 2.13. This figure clearly shows that, using the same interpolator, the reconstructed IMF slopes agree well with the input parameters.

For `MILES1` and `MILES3` we did a similar test. The results that we obtain are shown in Appendix 2.C. The slopes that we reconstruct by applying our model using the MILES templates of the same age and metallicity as the input SSP are also given in Table 2.7. For `MILES3` the reconstructed IMF slopes are now completely consistent with a Kroupa IMF. For `MILES1` we are able constrain the high-mass slope but the data does not allow us to reconstruct the low-mass slope. As one might expect, the results in Table 2.7 clearly show that it is easier to constrain the low-mass end of the IMF in older populations.

Our results show that the quality of the stellar templates that we provide to the model is a very important factor in our ability to reconstruct the IMF. If we use a set of stellar templates for which the systematic uncertainties are larger than the typical uncertainties in the data, we may find incorrect IMF parameters. The strength of our model lies in the fact that we can easily compare different sets of stellar templates based on their evidence. For example, if we apply our model to the MILES SSPs using the MILES templates, the log evidence is 28.3, 31.2, and 75.8 higher for, respectively, `MILES1`, `MILES2`, and `MILES3` than when we apply the model with our own stellar templates.

2.5.3 Mass fraction of low-mass stars

As in Section 2.4.2.3, we summarize our results by calculating the reconstructed mass fractions $F_{0.5}$. The reconstructed mass fractions for the MILES SSPs are given in Table 2.8. As a reference, this table also provides the mass fraction $F_{0.5}$ of the original IMF (Kroupa) for the same age and metallicity as the SSP. The results in Table 2.8 show that we are only able to reconstruct the true value of $F_{0.5}$ for the oldest SSP.

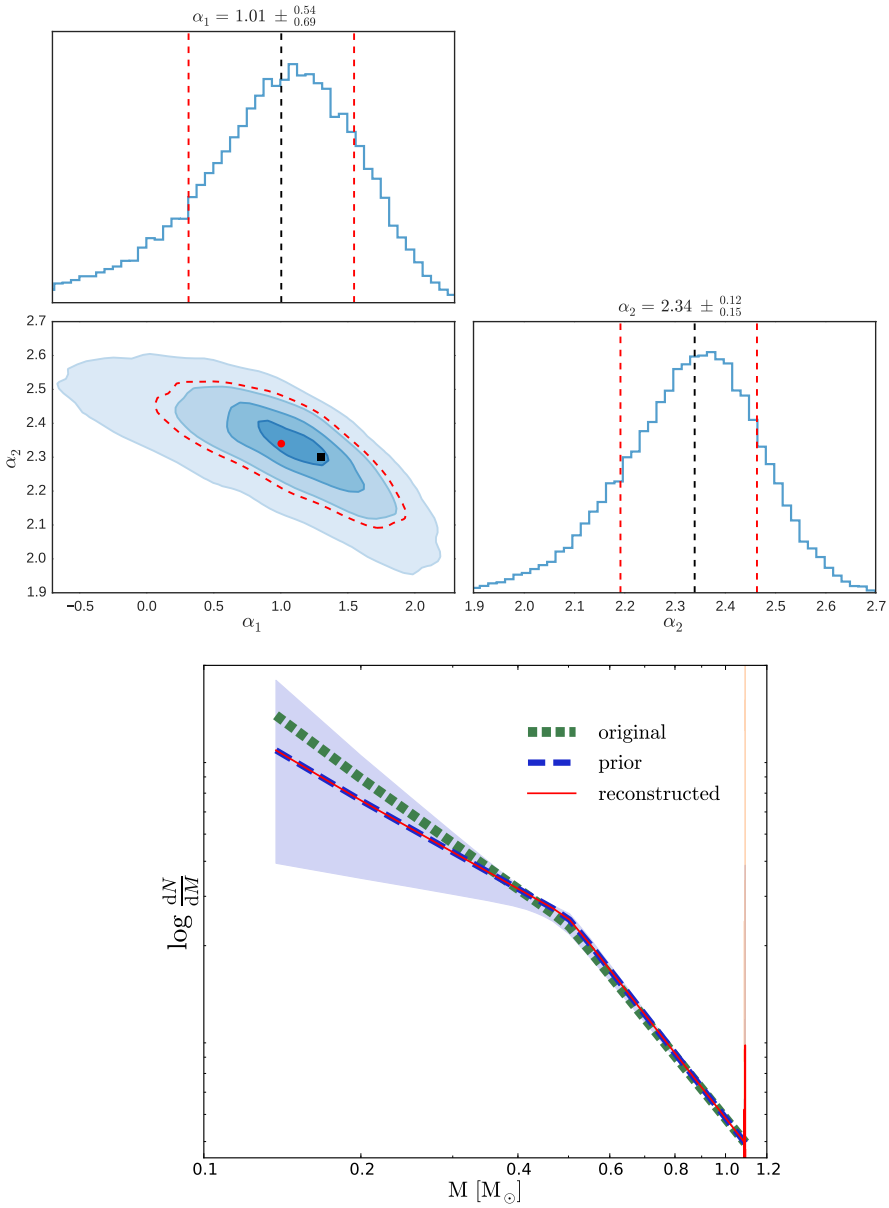

Figure 2.13 – As in Fig. 2.12, except for the age-metallicity plot in the top panel, but now using the stellar templates created with the interpolator on the MILES website.

Table 2.8 – Mass fraction of stars with $M < 0.5$ $M_\odot$ for the three MILES SSPs in Table 2.7. For each of the SSPs, the table provides both the mass fraction $F_{0.5}$ derived from the stellar templates described in Section 2.3 and the mass fraction $F_{0.5,\text{MILES templates}}$ derived from the MILES templates. As a reference, for every mock SSP we also provide $F_{0.5,\text{original}}$, the value of $F_{0.5}$ that the corresponding SSP would have for the input IMF (Kroupa).

name	$F_{0.5,\text{original}}$	$F_{0.5}$	$F_{0.5,\text{MILES templates}}$
MILES1	0.51	$0.76^{+0.01}_{-0.04}$	$0.41^{+0.06}_{-0.15}$
MILES2	0.58	$0.72^{+0.02}_{-0.07}$	$0.55^{+0.06}_{-0.09}$
MILES3	0.63	$0.67^{+0.03}_{-0.16}$	$0.67^{+0.02}_{-0.10}$

If we apply the MILES templates to the MILES SSPs, the reconstructed mass fractions $F_{0.5,\text{MILES templates}}$ agree better with the input data. Except for MILES1, for which the obtained mass fraction is slightly too low, the reconstructed mass fractions lie within one sigma of the mass fraction for a Kroupa IMF. Since we expect that approximately one third of the results is outside the one sigma contours, these results are consistent with the data.

The discrepancy between the results that we obtain for the two sets of stellar templates may be explained by the systematic differences between the stellar templates used to create the MILES SSPs and the stellar templates described in Section 2.3. These systematic differences have been discussed in Section 2.5.2 and include a different interpolator, different sets of isochrones and a different normalization scheme for the stellar templates. By using the stellar templates from the MILES website, we remove the interpolator as a source of systematic uncertainty. The remaining systematic uncertainties and the fact that the SNR of the low-mass stars is lower for younger populations (for a spectrum with constant SNR) most likely explains why the reconstructed value $F_{0.5,\text{MILES templates}}$ for MILES1 is inconsistent with the input IMF.

2.6 Summary and discussion

We have designed a new SPS code to reconstruct the shape of the IMF. The model that we have developed consists of a Bayesian framework with a number of different layers.

At the innermost level, the spectrum of a stellar population is represented as a linear combination of a set of stellar templates. For an SSP, these templates are defined by an isochrone. Combining an isochrone with a stellar library and an interpolator allows us to create a spectrum for each of these templates. The contribution of each of these spectra to the spectrum of the SSP (weights) is obtained through a linear inversion. We regularize this linear inversion by using a prior IMF, which translates into a prior on the weights.

The prior IMF that we use to regularize the linear inversion of the weights is chosen as being part of an IMF prior model family. This IMF prior model family is characterized by a set of (non-linear) parameters p_i. For every combination of p_i, we are able to construct a prior on the IMF which may be transformed into a prior on the weights. Given the input spectrum, the stellar templates, and the prior on the weights, our model determines the Bayesian evidence for that particular set of parameters. This allows us to sample the parameters of the IMF prior model family using MCMC techniques.

We then applied our model to a number of mock SSPs to demonstrate its validity. What we have shown is that we are able to reconstruct the input parameters of these mock SSPs. The quality of the reconstruction for these SSPs is mostly determined by the SNR of the input spectra and by the relative contribution of low-mass stars to the integrated spectrum. We have shown that the latter depends on both the age of the SSP and on the slope of the IMF. For younger SSPs, more light is emitted by stars more massive than 0.5 $M_\odot$. This effectively decreases the SNR of the stars less massive than 0.5 $M_\odot$. By increasing the low-mass slope of the IMF, we increase the number of low-mass stars which in turn increases the SNR of the low-mass stars. Constraining the low-mass IMF is therefore easier for older SSPs and for IMFs that are more bottom-heavy.

As a next step, we applied our model to three (mock) SSPs created with the MILES models. The age and metallicity reconstructed by our model are consistent with the input parameters. We are not able to correctly reconstruct the input IMF of the youngest MILES SSP. For the intermediate and oldest MILES SSPs there is also an offset between the input IMF and the reconstructed IMF. Nevertheless, for these two SSPs the input IMF is around the one sigma contour of the reconstructed IMF. The offsets that we find for the MILES SSPs may be explained by systematic differences between the MILES models and our models.

The application of our model to a set of mock SSPs shows that if the SNR of a spectrum is high enough to reveal the signal of the low-mass stars, in principle we are able to reconstruct the IMF of these SSPs. However, the application of the model to three MILES SSPs demonstrates that if there are systematic uncertainties this may introduce a bias on the obtained results. At the moment we do not take these systematic uncertainties into account but the results for the MILES SSPs show that it is crucial to model these uncertainties as well.

One of the most important sources of systematic uncertainty between the MILES models and our model is the interpolator that is used to create the stellar templates. To demonstrate this, we have downloaded a set of interpolated spectra from the MILES website using their interpolator. Using these spectra, we are able to reconstruct the input IMF of the MILES SSPs with the exception of the low-mass slope of the youngest SSP. This demonstrates very clearly that our model is only able to reconstruct the IMF if we provide it with a representative set of stellar templates. In that respect, the Bayesian framework of our model is very important as it allows us to objectively compare different model ingredients with each other in light of the evidence.

The application of our model to the MILES models shows us that the reconstructed IMF slopes are very sensitive to the interpolation method used. Therefore, reconstructing the IMF requires a reliable interpolator. In practice, interpolation between stellar spectra can be difficult. The stellar libraries on which interpolators are based, in general, do not provide complete coverage of the parameter space, and there is uncertainty in the parameters of the stars in the library. Moreover, one has to define the parameters that are relevant for interpolating between stellar spectra. For now, we use the effective temperature, surface gravity and [Fe/H] ratio. However, for some stellar populations it may be necessary to also include, for example, the [α/Fe] ratio. In addition to that, there are variable stars and stars with peculiarities. These stars are very hard to model and most often found in the low-mass end of the main sequence and the upper giant and asymptotic giant branch.

One of the main questions that we are trying to answer by reconstructing the IMF is the ratio of dwarf to giant stars. This question is difficult to answer in the spectral window of the MILES library because the great spectral similarity between low-mass stars and K and M giants makes these objects difficult to differentiate. In the future we plan to use the X-shooter

Spectral Library (XSL) as an input to our model. This stellar library offers much broader wavelength coverage, extending from the UV to the NIR. Compared to MILES, the spectral window of XSL contains many more IMF-sensitive features that will help to better constrain the IMF of distant galaxies.

Acknowledgements

This research is partially based on data from the MILES project. We thank Philippe Prugniel and Alexandre Vazdekis for helpful discussions. We thank the anonymous referee for useful comments which improved the quality of this paper. This work was supported in part by an NWO grant (project number 614.001.208) to SCT and by an NWO-VICI career grant (project number 639.043.308) to LVEK.

Appendix 2.A Velocity dispersion mock SSPs

If we increase the velocity dispersion of the mock SSPs, more and more spectral features will be washed out of the spectrum. For very high velocity dispersion, one can therefore imagine that this becomes problematic for reconstructing the IMF. However, if the velocity dispersion is not too high we do not expect that our results depend on the velocity dispersion of the mock SSP.

To test the robustness of our results as a function of velocity dispersion we reconsider mock SSP `mock5`. We create five different versions of this mock SSP, each of them smoothed to a different velocity dispersion (but using the same noise spectrum). For each of these mock SSPs we reconstruct the low-mass slope α_1 and the high-mass slope α_2. The different velocity dispersions and the results for the reconstruction of α_1 and α_2 are given in Table 2.9. Over a range in velocity dispersion of $0 - 300\,\mathrm{km\,s^{-1}}$, the inferred IMF parameters show no dependence on velocity dispersion. Hence, we do not expect that our choice to smooth our spectra to a velocity dispersion of 150 km s^{-1} has implications for the reconstructed IMFs that we obtain.

Table 2.9 – Reconstruction of low-mass slope α_1 and high-mass slope α_2 for `mock5` as a function of velocity dispersion.

velocity dispersion [km s^{-1}]	α_1	α_2
0	1.26 ± 0.51	2.23 ± 0.16
75	1.26 ± 0.53	2.23 ± 0.16
150	1.26 ± 0.51	2.23 ± 0.17
225	1.27 ± 0.56	2.23 ± 0.18
300	1.29 ± 0.48	2.21 ± 0.16

Appendix 2.B Results for mock SSPs

In Section 2.4.2.2 we showed the results for the reconstruction of the IMF for a mock SSP with an age of 8.5 Gyr and metallicity [Fe/H] = 0.0. The plots in this appendix show the results that we obtain for the remaining mock SSPs discussed in Section 2.4.2.2. Table 2.1 gives an overview of the twelve mock SSPs that we consider.

For each of the remaining mock SSPs we show the distribution of the evidence in the age-metallicity grid, the reconstruction of the (non-linear) IMF slopes and the (linear) reconstruction of the piecewise IMF by using the best-fitting non-linear parameters as a prior.

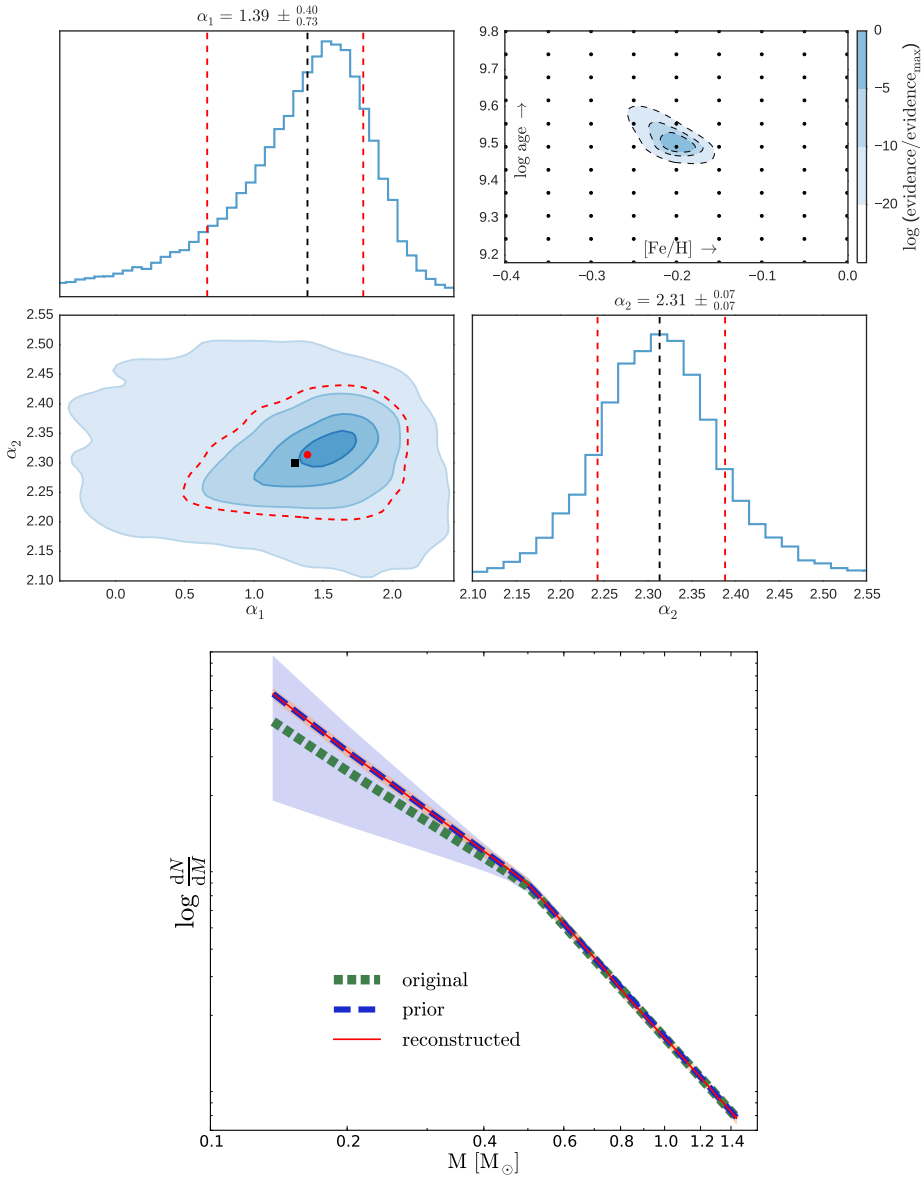

Figure 2.14 – Results for `mock1` with $t = 3.1$ Gyr, [Fe/H] $= -0.2$ and a Kroupa IMF. **Top panels:** As in Fig. 2.8 only now for `mock1`. **Bottom panel:** As in the top panel of Fig. 2.9 only now for `mock1`

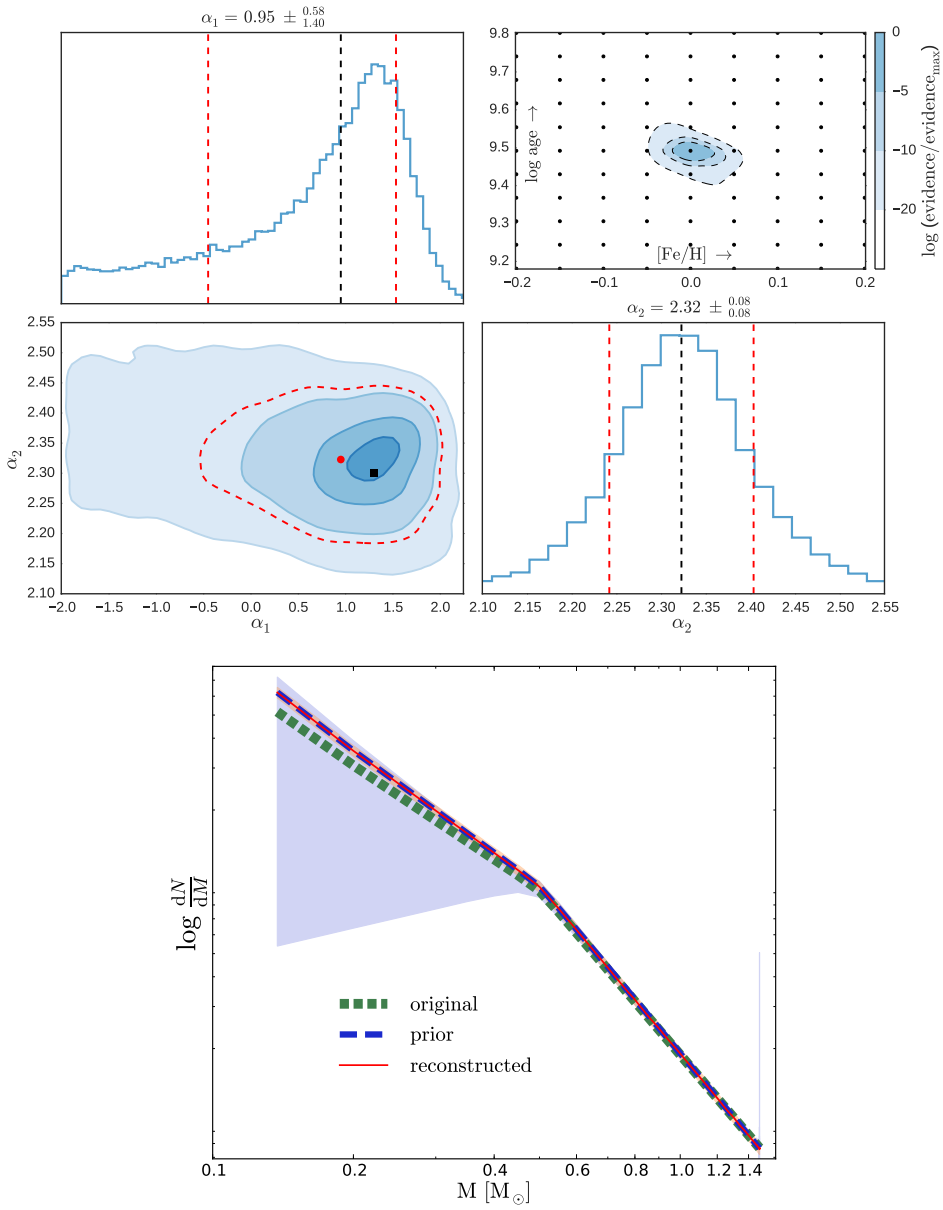

Figure 2.15 – As in Fig. 2.14 for `mock2` with $t = 3.1$ Gyr, [Fe/H] $= 0.0$ and a Kroupa IMF.

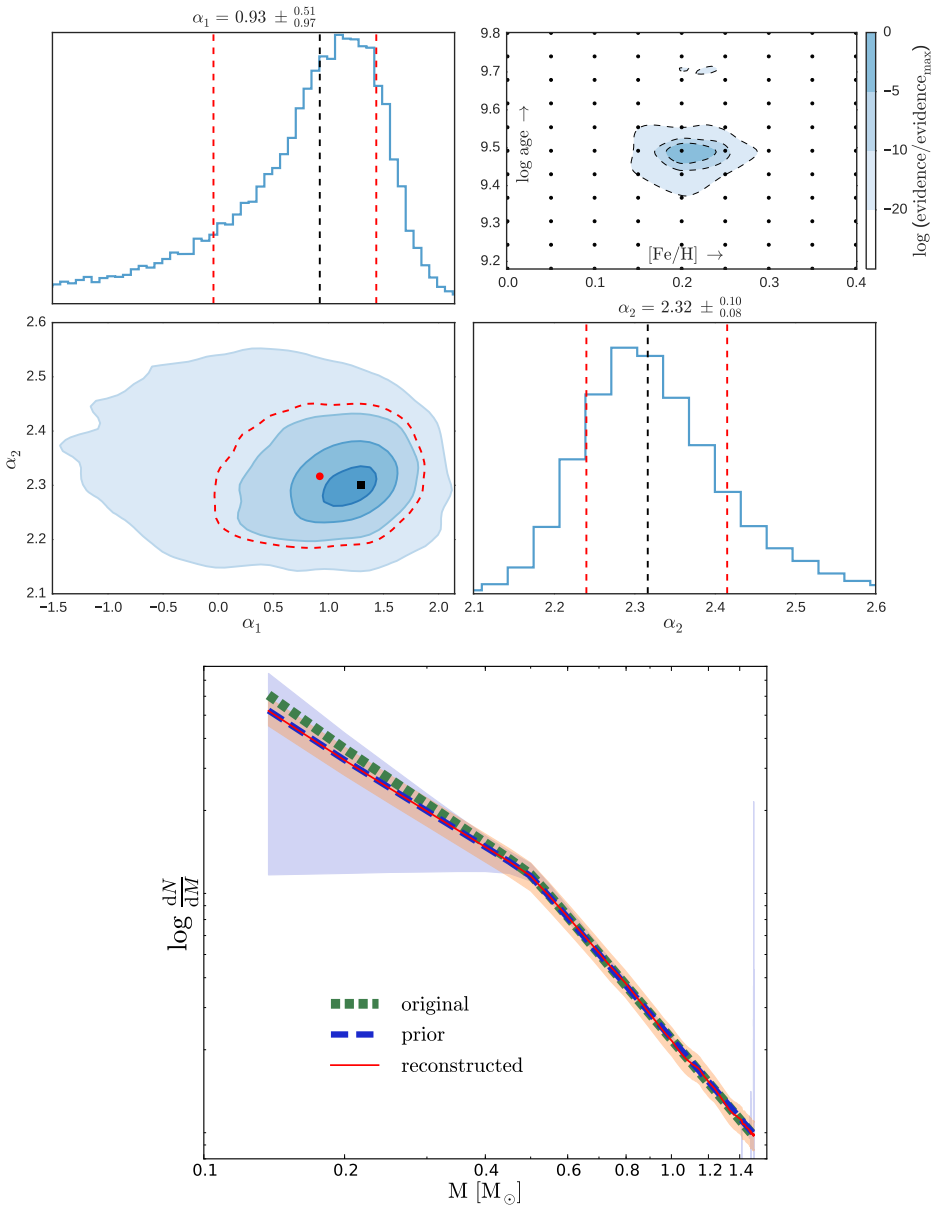

Figure 2.16 – As in Fig. 2.14 for `mock3` with $t = 3.1$ Gyr, [Fe/H] $= 0.2$ and a Kroupa IMF.

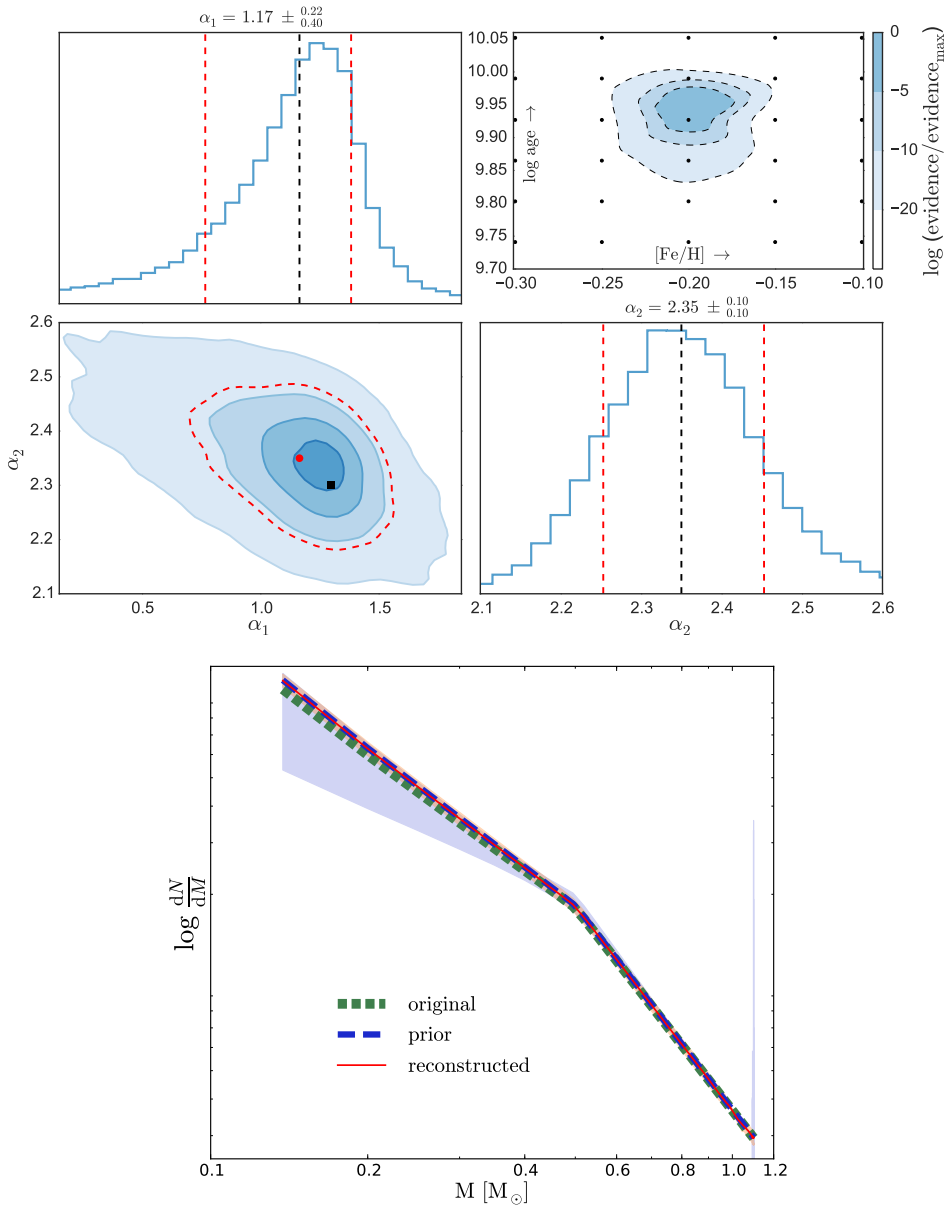

Figure 2.17 – As in Fig. 2.14 for `mock4` with $t = 8.5$ Gyr, [Fe/H] $= -0.2$ and a Kroupa IMF.

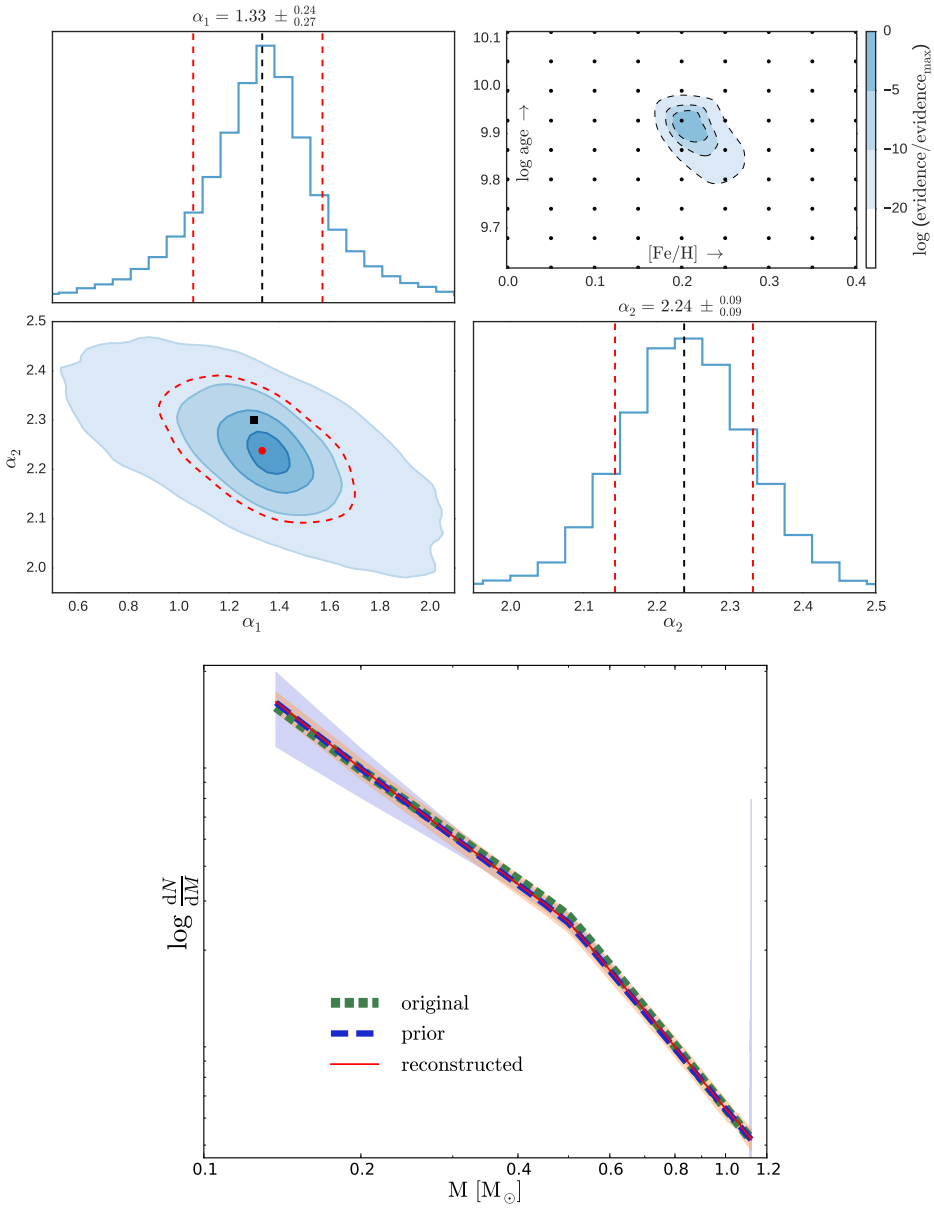

Figure 2.18 – As in Fig. 2.14 for `mock6` with $t = 8.5$ Gyr, [Fe/H] $= 0.2$ and a Kroupa IMF.

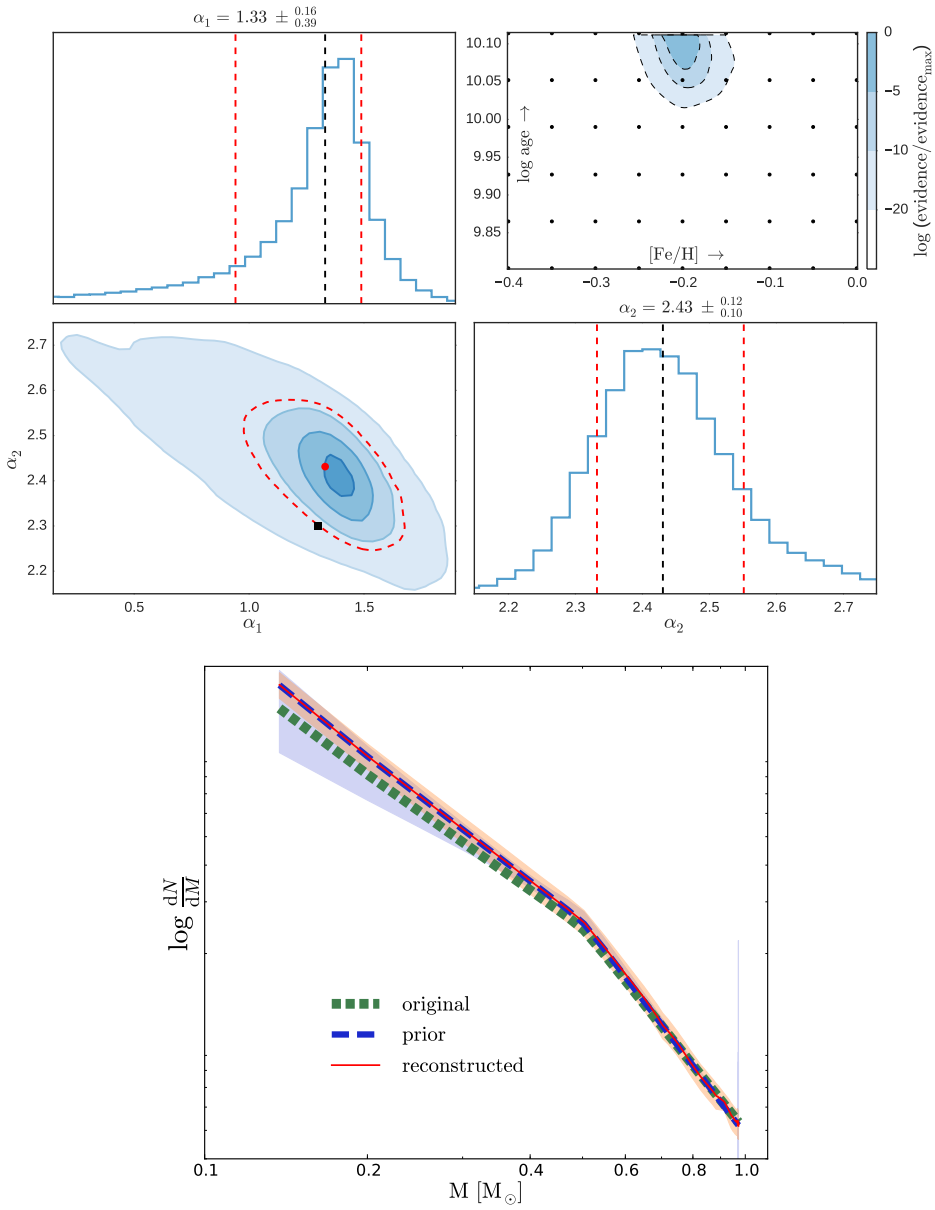

Figure 2.19 – As in Fig. 2.14 for `mock7` with $t = 13.0$ Gyr, [Fe/H] $= -0.2$ and a Kroupa IMF.

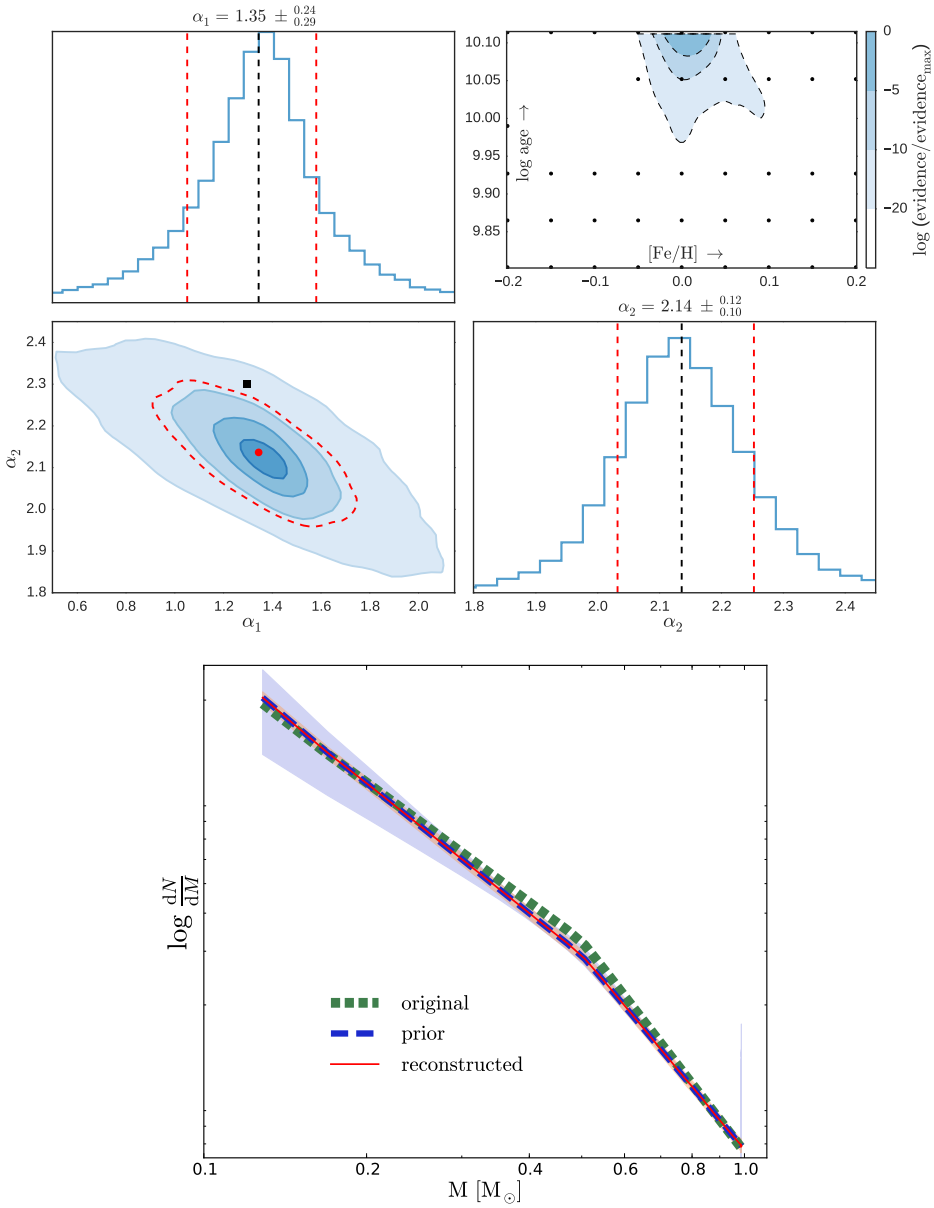

Figure 2.20 – As in Fig. 2.14 for `mock8` with $t = 13.0$ Gyr, [Fe/H] $= 0.0$ and a Kroupa IMF.

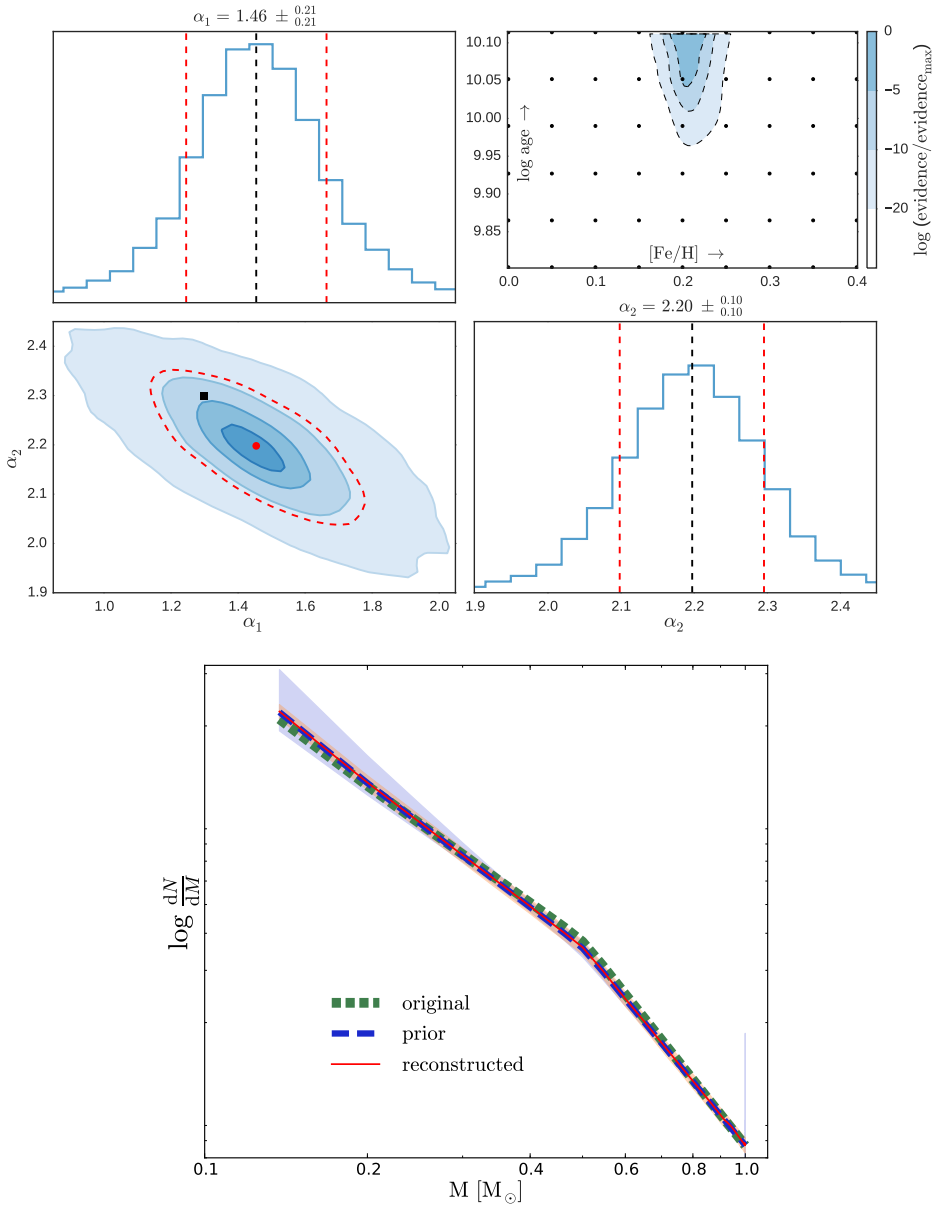

Figure 2.21 – As in Fig. 2.14 for `mock9` with $t = 13.0$ Gyr, [Fe/H] $= 0.2$ and a Kroupa IMF.

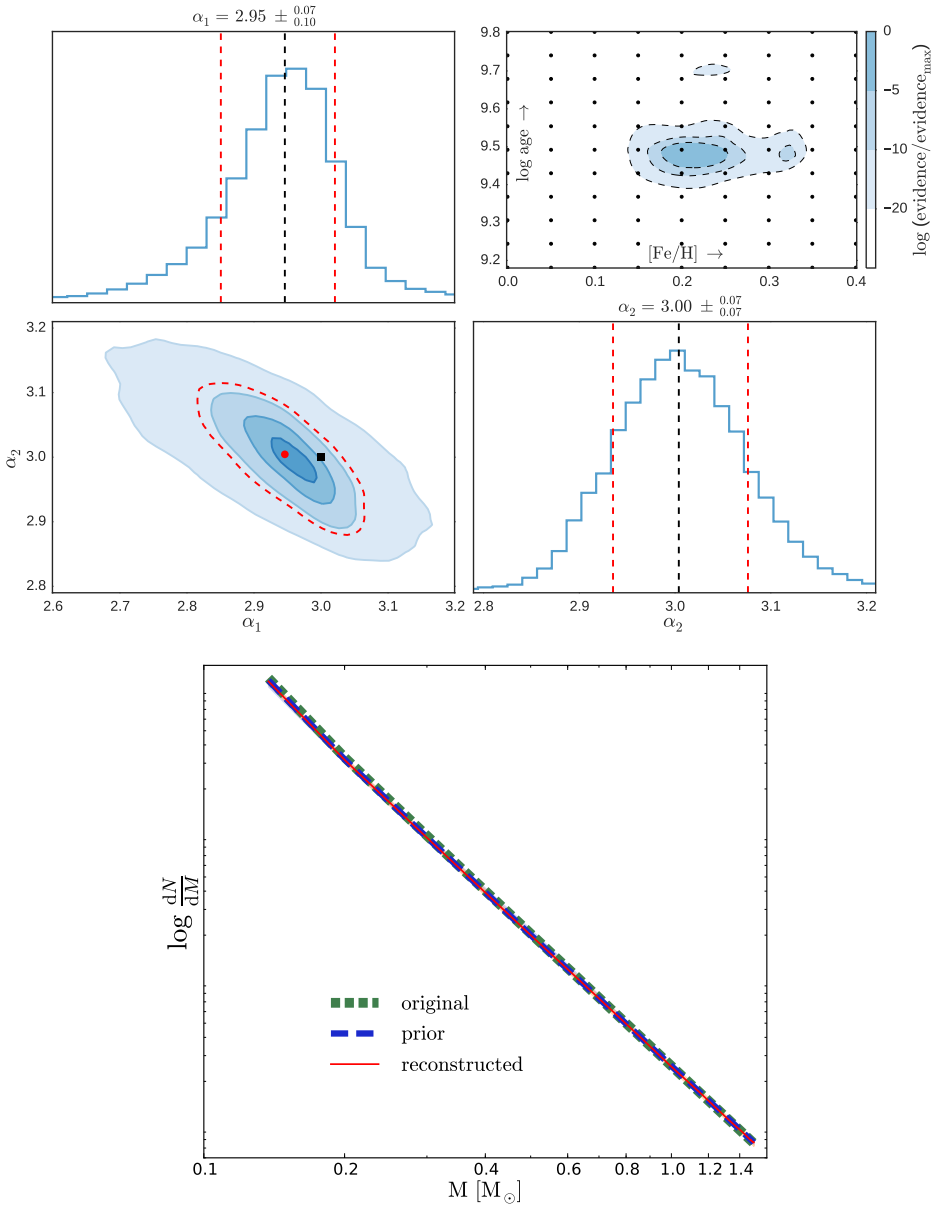

Figure 2.22 – As in Fig. 2.14 for `mock10` with $t = 3.1$ Gyr, [Fe/H] $= 0.2$ and a bottom-heavy IMF.

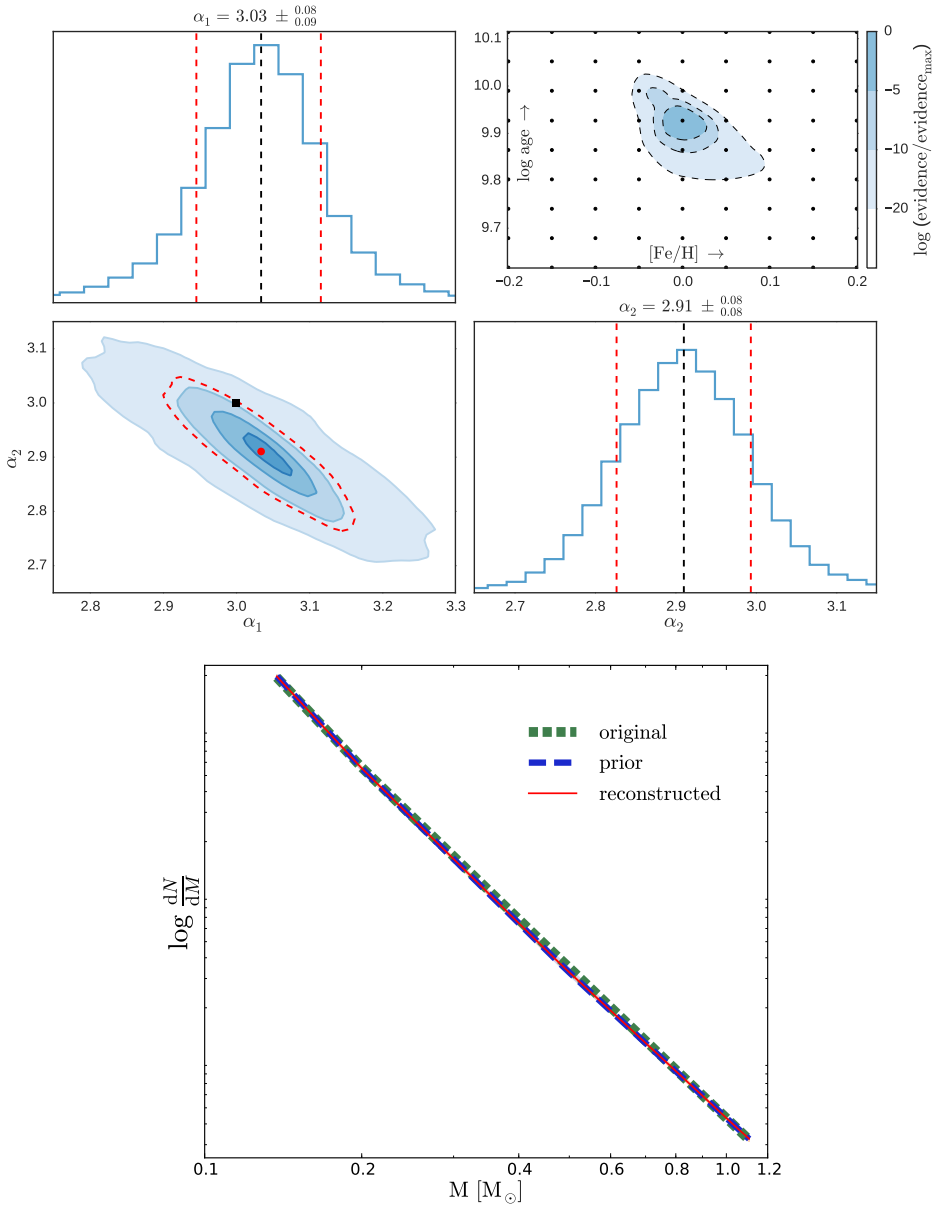

Figure 2.23 – As in Fig. 2.14 for `mock11` with $t = 8.5$ Gyr, [Fe/H] $= 0.0$ and a bottom-heavy IMF.

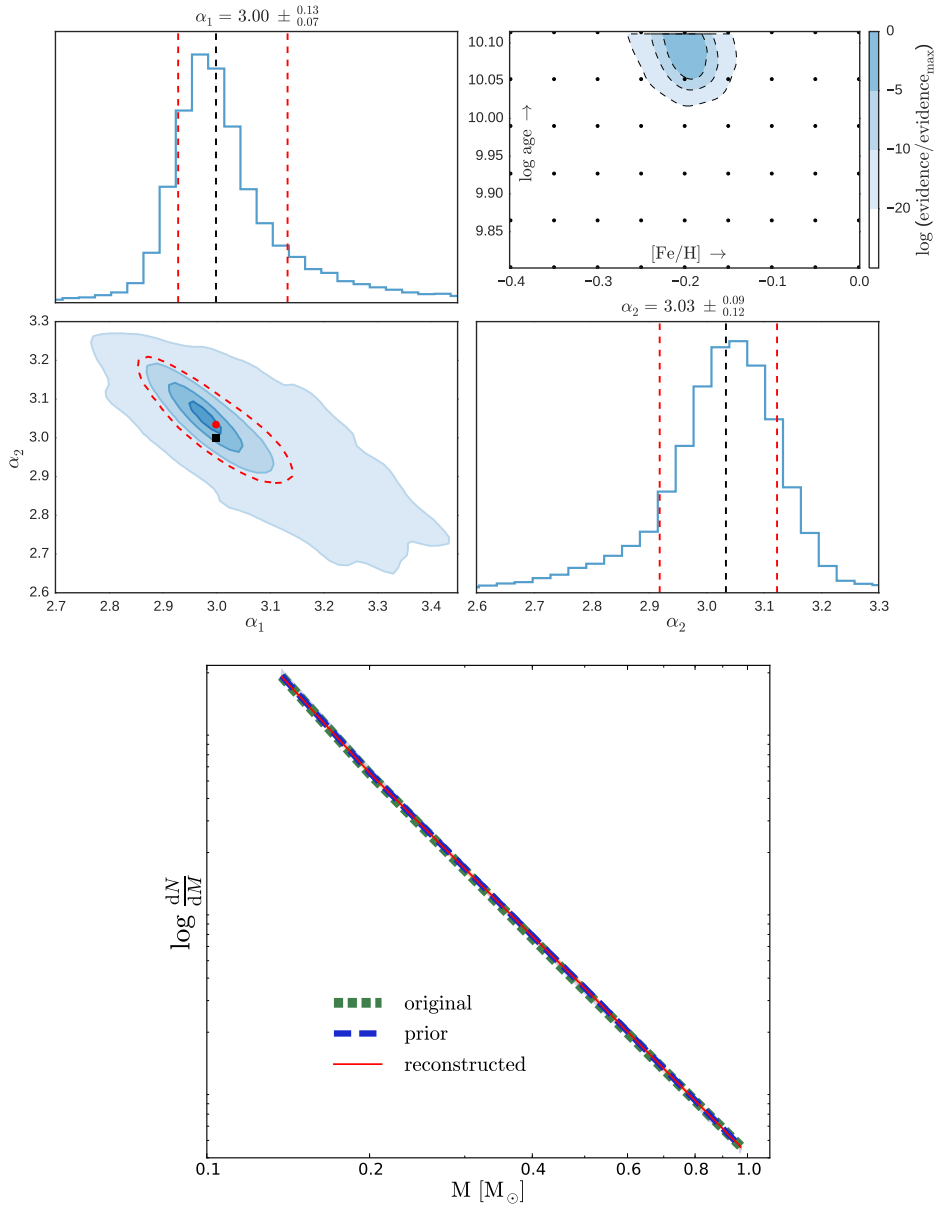

Figure 2.24 – As in Fig. 2.14 for `mock12` with $t = 13.0$ Gyr, [Fe/H] $= -0.2$ and a bottom-heavy IMF.

Appendix 2.C Results for MILES SSPs

In Section 2.5 we have discussed the application of our model to three MILES SSPs. The results for `MILES2` have been shown in Section 2.5. Here we show the results for `MILES1` and `MILES3`

The reconstructed IMF for `MILES1` (Fig. 2.25) suggests a very steep low-mass IMF combined with a high-mass slope that is much flatter than it is for a Kroupa IMF. This result is inconsistent with the input IMF. The over-abundance of the lowest-mass templates ($M \lesssim 0.3$ M$_\odot$) with respect to the input Kroupa IMF produces a relatively small amount of light. Therefore, the under-abundance of the intermediate-mass templates (0.3 M$_\odot \lesssim M \lesssim 1.2$ M$_\odot$) is almost completely compensated by a slight over-abundance of the highest mass templates ($M \gtrsim 1.2$ M$_\odot$).

The reconstructed IMFs for the MILES SSPs show a spiky distribution for the high-mass end. There are three reasons that might explain this behaviour. First, at the high-mass end the difference between two subsequent mass bins may be very small. To convert the reconstructed weights into an IMF, we divide the weights by the width of the mass bin. As a consequence, a relatively small deviation of the weights may result in a large deviation of the IMF if the mass bin is small. Second, the regularization scheme that we use for the MILES SSPs penalizes the absolute deviation of the weights. This implies that a deviation of one star from the prior for a low-mass template is penalized as much as a deviation of one star for a high-mass template. However, since the number of stars for higher-mass templates is in general much lower (because of the shape of the IMF and the width of the mass bins) the effect on the IMF will be much larger. Third, the use of NNLS may result in some of the templates being equal to zero. If this is the case, the model has to compensate for this by increasing the weights of the templates that surround the one that is set to zero in the fit.

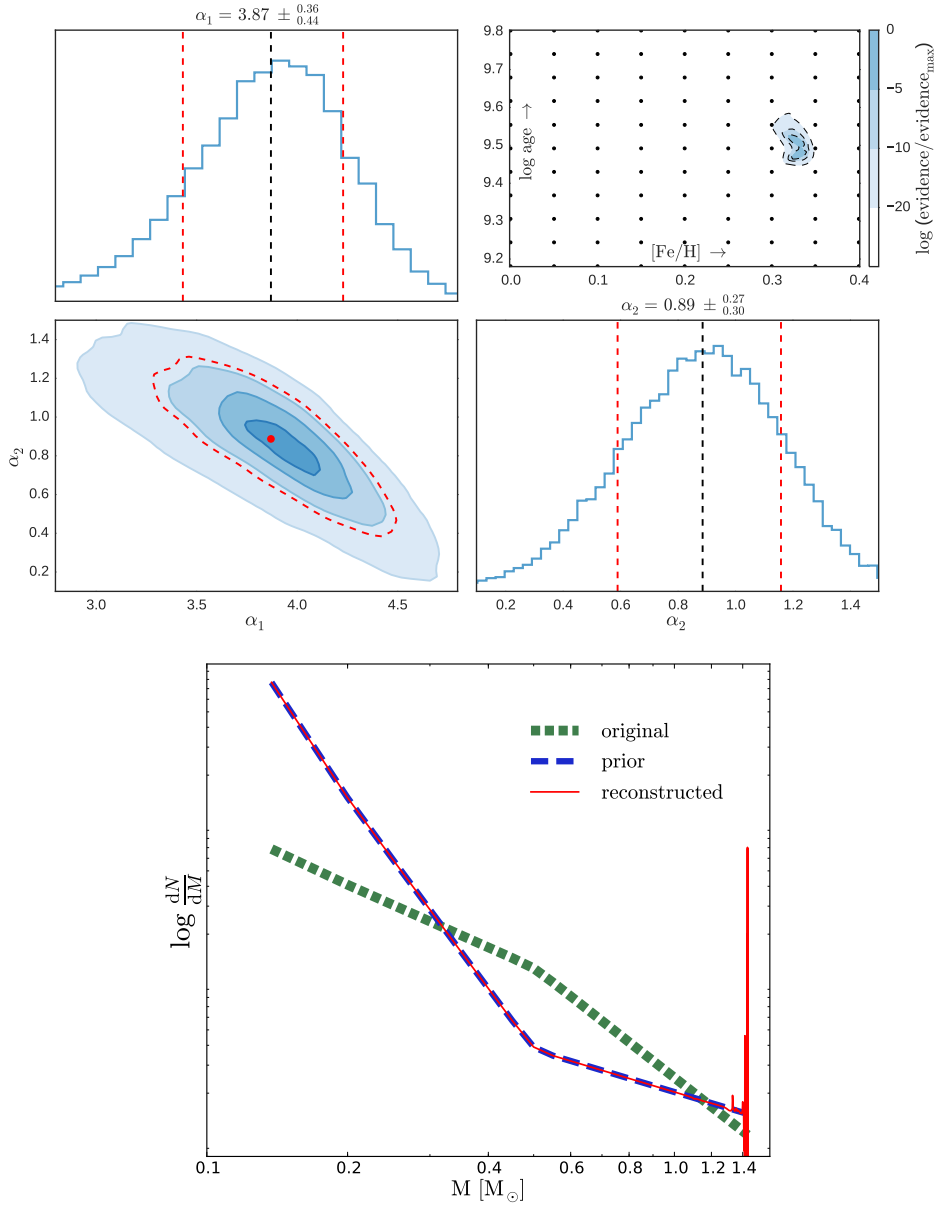

Figure 2.25 – As in Fig. 2.14 for `MILES1` with $t = 3.2$ Gyr and [Fe/H] $= 0.22$.

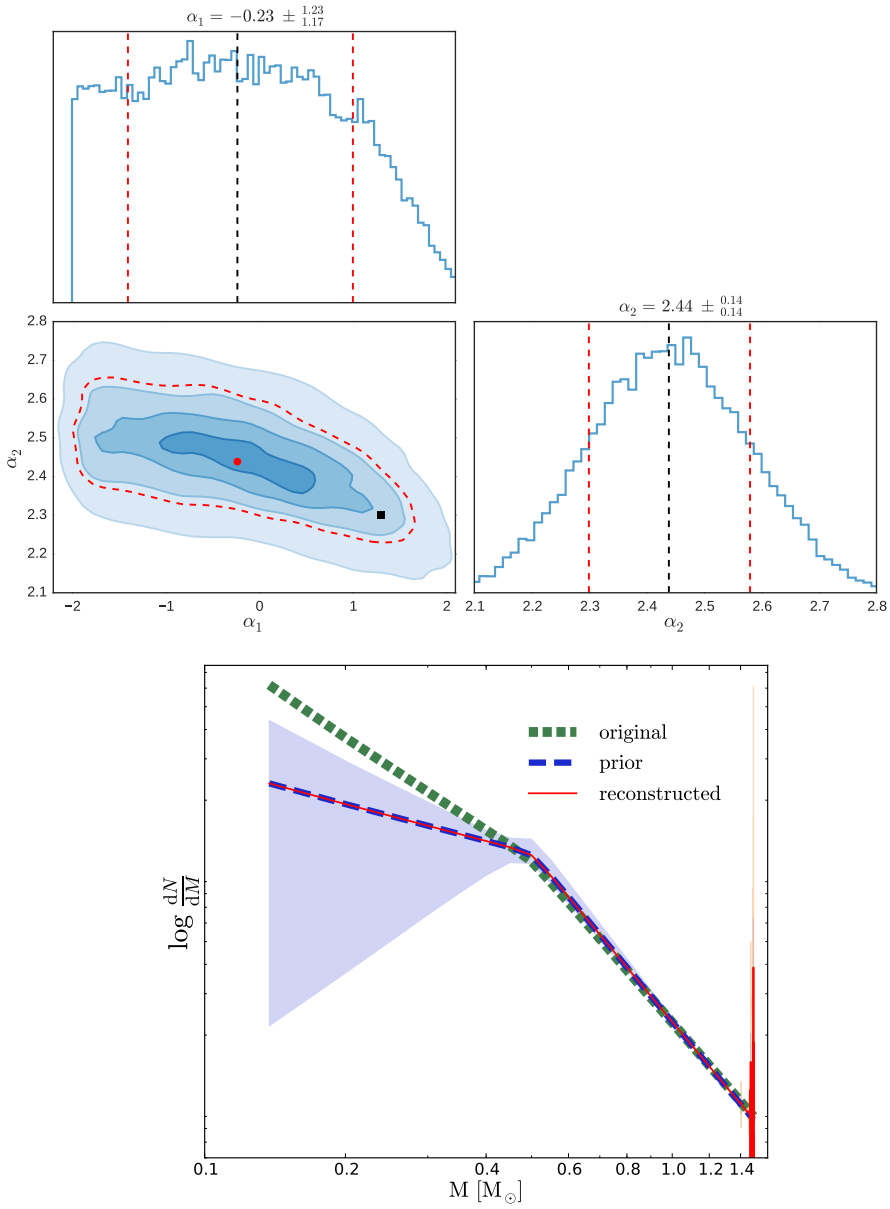

Figure 2.26 – As in Fig. 2.25 for `MILES1` with $t = 3.2$ Gyr and [Fe/H] $= 0.22$ but now using the stellar templates created using the interpolator on the MILES website.

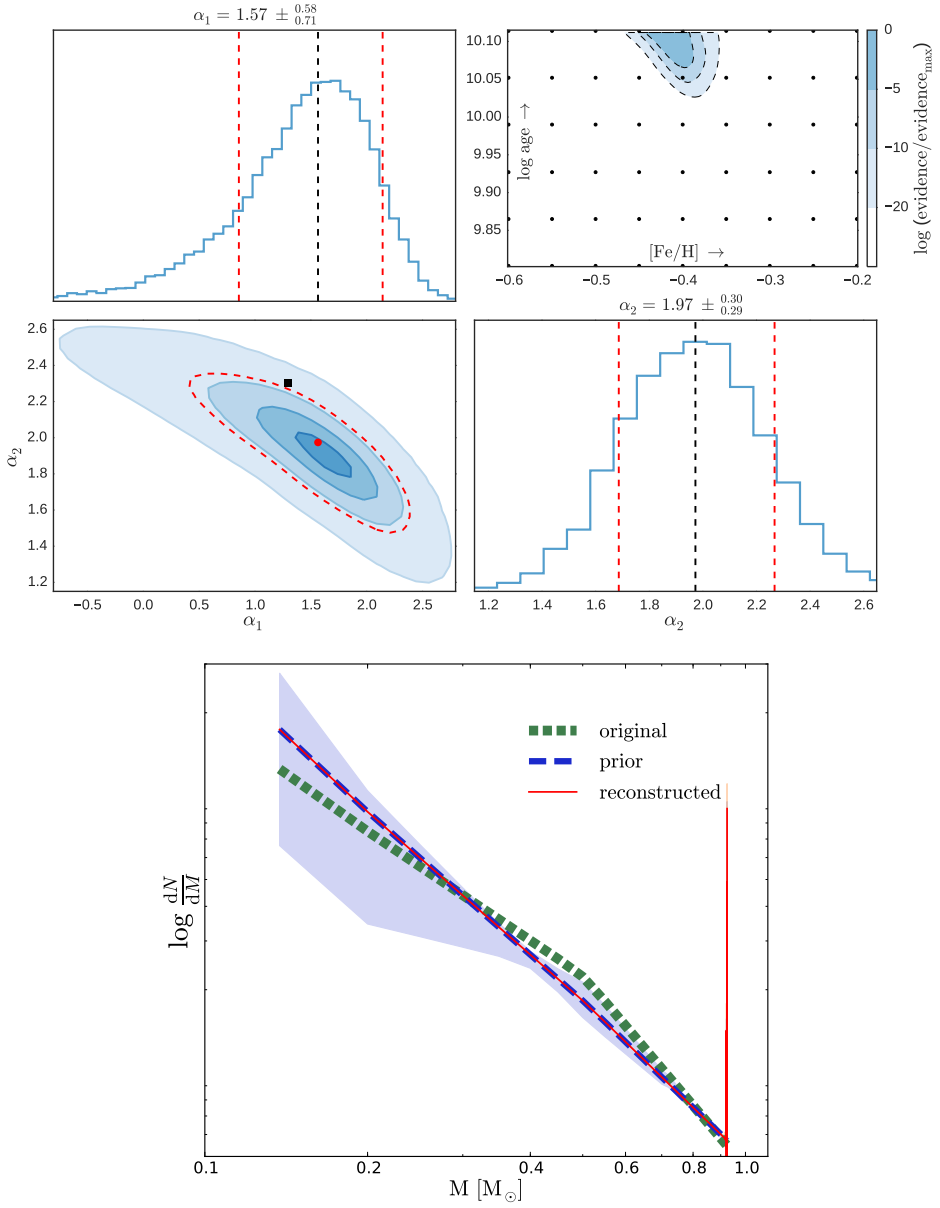

Figure 2.27 – As in Fig. 2.14 for `MILES3` with $t = 12.6$ Gyr and [Fe/H] $= -0.40$.

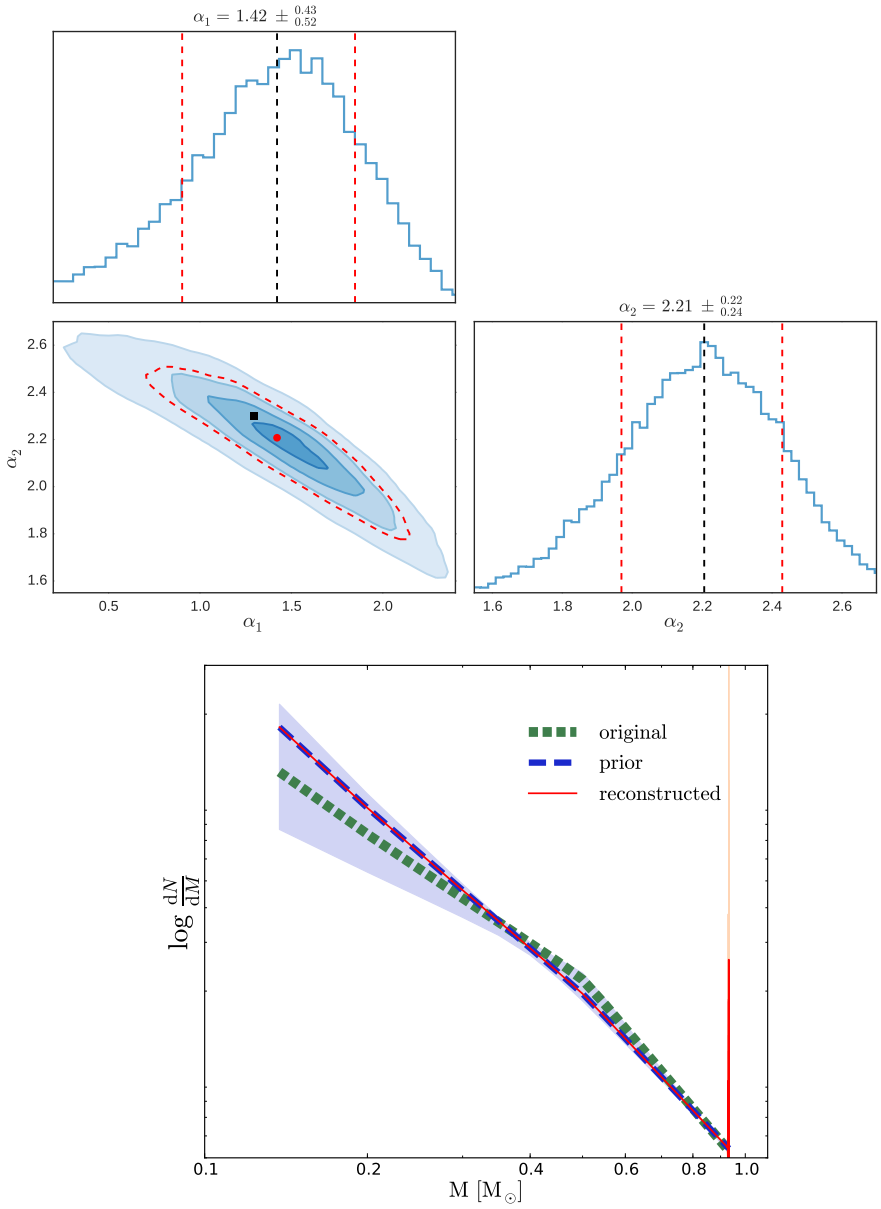

Figure 2.28 – As in Fig. 2.27 for `MILES3` with $t = 12.6$ Gyr and [Fe/H] $= -0.40$ but now using the stellar templates created using the interpolator on the MILES website.

References

Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000
Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2012, Nature, 484, 485
Cenarro, A. J., Gorgas, J., Vazdekis, A., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 339, L12
Cenarro, A. J., Peletier, R. F., Sánchez-Blázquez, P., et al. 2007, MNRAS, 374, 664
Chabrier, G. 2003, PASP, 115, 763
Chen, Y.-P., Trager, S. C., Peletier, R. F., et al. 2014, A&A, 565, A117
Conroy, C. & van Dokkum, P. 2012, ApJ, 747, 69
Dabringhausen, J., Hilker, M., & Kroupa, P. 2008, MNRAS, 386, 864
Davé, R. 2008, MNRAS, 385, 147
Faber, S. M. & French, H. B. 1980, ApJ, 235, 405
Feroz, F. & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, ArXiv e-prints
Ferreras, I., La Barbera, F., de la Rosa, I. G., et al. 2013, MNRAS, 429, L15
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gorgas, J., Faber, S. M., Burstein, D., et al. 1993, ApJS, 86, 153
Graves, G. J. & Faber, S. M. 2010, ApJ, 717, 803
Jeffreys, H. 1961, Oxford Univ. Press
Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, A&A, 501, 1269
Kroupa, P. 2001, MNRAS, 322, 231
Kroupa, P., Tout, C. A., & Gilmore, G. 1993, MNRAS, 262, 545
La Barbera, F., Ferreras, I., Vazdekis, A., et al. 2013, MNRAS, 433, 3017
Lawson, C. & Hanson, R. 1995, SIAM
Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. 2004, A&A, 425, 881
MacKay, D. J. C. 1992, Neural Computation, 4, 415
Maíz Apellániz, J. 2006, AJ, 131, 1184
Maraston, C. 2005, MNRAS, 362, 799
Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883
Miller, G. E. & Scalo, J. M. 1979, ApJS, 41, 513

Prugniel, P., Vauglin, I., & Koleva, M. 2011, A&A, 531, A165
Salpeter, E. E. 1955, ApJ, 121, 161
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Schiavon, R. P. 2007, ApJS, 171, 146
Schiavon, R. P., Barbuy, B., Rossi, S. C. F., & Milone, A. 1997a, ApJ, 479, 902
Schiavon, R. P., Barbuy, B., & Singh, P. D. 1997b, ApJ, 484, 499
Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, American Institute of Physics Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405
Spiniello, C., Trager, S., Koopmans, L. V. E., & Conroy, C. 2014, MNRAS, 438, 1483
Spiniello, C., Trager, S. C., Koopmans, L. V. E., & Chen, Y. P. 2012, ApJL, 753, L32
Tinsley, B. M. 1968, ApJ, 151, 547
Tinsley, B. M. 1972, ApJ, 178, 319
Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, ApJ, 709, 1195
van Dokkum, P. G. 2008, ApJ, 674, 29
Vazdekis, A., Cenarro, A. J., Gorgas, J., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 340, 1317
Vazdekis, A., Coelho, P., Cassisi, S., et al. 2015, MNRAS, 449, 1177
Vazdekis, A., Ricciardelli, E., Cenarro, A. J., et al. 2012, MNRAS, 424, 157
Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, MNRAS, 404, 1639
Wing, R. F. & Ford, Jr., W. K. 1969, PASP, 81, 527
Worthey, G., Faber, S. M., Gonzalez, J. J., & Burstein, D. 1994, ApJS, 94, 687
Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45

Chapter 3

Hierarchical Bayesian inference of the Initial Mass Function in Composite Stellar Populations

— **M. Dries, S.C. Trager, L.V.E. Koopmans, G. Popping and R.S. Somerville** —

MNRAS, 2018, 474, 3500

Abstract

The initial mass function (IMF) is a key ingredient in many studies of galaxy formation and evolution. Although the IMF is often assumed to be universal, there is continuing evidence that it is not universal. Spectroscopic studies that derive the IMF of the unresolved stellar populations of a galaxy often assume that this spectrum can be described by a single stellar population (SSP). To alleviate these limitations, in this paper we have developed a unique hierarchical Bayesian framework for modelling composite stellar populations (CSPs). Within this framework we use a parameterized IMF prior to regulate a direct inference of the IMF. We use this new framework to determine the number of SSPs that is required to fit a set of realistic CSP mock spectra. The CSP mock spectra that we use are based on semi-analytic models and have an IMF that varies as a function of stellar velocity dispersion of the galaxy. Our results suggest that using a single SSP biases the determination of the IMF slope to a higher value than the true slope, although the trend with stellar velocity dispersion is overall recovered. If we include more SSPs in the fit, the Bayesian evidence increases significantly and the inferred IMF slopes of our mock spectra converge, within the errors, to their true values. Most of the bias is already removed by using two SSPs instead of one. We show that we can reconstruct the variable IMF of our mock spectra for signal-to-noise ratios exceeding $\sim$75.

3.1 Introduction

The distribution of stellar masses in a stellar population or a galaxy is described by the current mass function (CMF). The CMF is related to the initial mass function (IMF) which describes the distribution of stellar masses at the time that the stars were born. The concept of an IMF was first introduced by Salpeter (1955) who called it the 'original mass function'. Based on star counts of resolved stellar populations, the Milky Way (MW) IMF can be described by a broken power law (Kroupa et al. 1993; Kroupa 2001) or a lognormal distribution extended with a power law for higher masses (Chabrier 2003). Most studies of the MW's IMF in different environments suggest a universal IMF (Bastian et al. 2010).

More distant galaxies are unresolved and therefore the IMF cannot be determined directly from star counts. Whereas the Galactic IMF is determined for a spiral galaxy, unresolved stellar population studies often focus on early-type (elliptical and lenticular) galaxies, which may not form stars following the MW's IMF. For those galaxies, the IMF may be inferred by modelling their spectra with a stellar population synthesis (SPS) model (e.g. Spinrad & Taylor 1971; Cenarro et al. 2003; Conroy & van Dokkum 2012a). Recent spectroscopic studies of early-type galaxies (ETGs) suggest that the IMF is not universal and that the relative number of low-mass ($M \lesssim 0.5$ $M_\odot$) to high-mass stars is higher in galaxies with a higher mass / velocity dispersion (Conroy & van Dokkum 2012b; Spiniello et al. 2012; Ferreras et al. 2013; La Barbera et al. 2013) or metallicity (Martín-Navarro et al. 2015b). Within these ETGs there appears to be a radial trend such that the IMF is more bottom-heavy towards the centre (Martín-Navarro et al. 2015a; van Dokkum et al. 2016). Combined studies of population synthesis with dynamics and/or gravitational lensing support that these ETGs have more stellar mass than predicted on the basis of a MW IMF (Treu et al. 2010; Graves & Faber 2010; Cappellari et al. 2012; Conroy & van Dokkum 2012b; Lyubenova et al. 2016). In contrast to these findings, Smith et al. (2015) and Newman et al. (2016) have found a number of lensed ETGs that favour a Kroupa IMF instead of a bottom-heavy IMF.

Inferring the IMF of a galaxy from its spectrum through a SPS model is not straightforward, however. Since Tinsley (1968), many SPS models (e.g. Bruzual & Charlot 2003; Le Borgne et al. 2004; Maraston 2005; Conroy & van Dokkum 2012a; Vazdekis et al. 2015) have been developed to infer the properties of a galaxy that are not directly observable. Every

SPS model has its own set of ingredients (basically stellar evolution in the form of isochrones, a spectral library and an IMF supplemented with an arbitrary number of model-specific ingredients/parameters and model assumptions) and each of these ingredients is characterized by uncertainties. The ingredients that are or are not included as well as their uncertainties might affect the inferred IMF.

One very noticeable example of such an 'ingredient' is that most SPS models assume a parameterized IMF and therefore do not try to determine the IMF shape but simply determine one or two IMF slopes. This in turn controls the ratio between dwarf and giant stars. Clauwens et al. (2016) show that different parameterizations of the IMF can results in similar dwarf-to-giant ratios. More direct approaches to determine the IMF shape have been developed by Dries et al. (2016) and Conroy et al. (2017). These methods try to determine the contribution of individual mass bins to the integrated spectrum and use this to infer the IMF.

Dwarfs with $M < 0.5$ $M_\odot$ contribute significantly to the stellar mass of a galaxy. The fraction of the stellar mass in stars with $M < 0.5$ $M_\odot$ is $\sim$27% for a Kroupa IMF between 0.1 $M_\odot$ and 100 $M_\odot$[1]. For a Salpeter IMF this fraction increases to $\sim$47% and for IMFs more bottom-heavy than Salpeter this fraction is even higher. However, the relative contribution of these dwarfs to an optical spectrum is expected to be not more than a few percent for a MW IMF (Dries et al. 2016, DTK16 hereafter). Moreover, the spectra of these dwarfs and the spectra of K and M giants look very similar and this makes it difficult to determine dwarf-to-giant ratios. There are, however, some (gravity-sensitive) spectral features that are known to be sensitive to either dwarfs or giants (Wing & Ford 1969; Faber & French 1980; Gorgas et al. 1993; Worthey et al. 1994; Schiavon et al. 1997a,b; Cenarro et al. 2003; Schiavon 2007; Spiniello et al. 2012). A particular example of such a feature is the CaH1 index defined in Spiniello et al. (2014). This feature is relatively strong in M-dwarfs but almost absent in M-giants (Spiniello et al. 2014).

Many SPS models fit the spectra of ETGs by assuming one or two single stellar populations (SSPs). In this work we test to what extent this assumption is justified by applying an improved version of the model by DTK16 to a set of realistic composite stellar population (CSP) mock

[1]In Dries et al. (2016) the fraction of mass in stars with $M < 0.4$ $M_\odot$ is erroneously reported to be 12% for a Kroupa IMF between 0.1 and 100 $M_\odot$. The correct number should be 21%.

spectra. We use the Bayesian evidence to test how many SSPs are typically necessary to describe a given CSP. Assuming an IMF that varies with velocity dispersion according to Spiniello et al. (2014), we investigate if we can recover this velocity dispersion–IMF relation from our mock spectra. The outline is as follows. In Section 2 we describe how we extend the DTK16 SSP model to a fully Bayesian CSP model, which uses the evidence for model comparison. In Section 3 we describe the construction of our stellar templates. We specify the CSP mock spectra in Section 4. In Section 5 we report our results and conclusions.

3.2 Model description

In this work we use the hierarchical Bayesian framework developed by DTK16 for reconstructing the IMF of SSPs. Here, we extend that work with a number of new features that allow us to model the IMF of CSPs.

3.2.1 Hierarchical Bayesian framework

The basic assumption of the hierarchical Bayesian framework developed by DTK16 is that the spectrum of a galaxy can be written as the sum of the spectra of all the stars that it contains. This is expressed by the following matrix equation:

$$\boldsymbol{g} = \mathbf{S}\,\boldsymbol{w}, \tag{3.1}$$

in which $\boldsymbol{g}$ is a column vector containing the intensity as a function of wavelength for a population of stars (e.g. a galaxy), $\mathbf{S}$ is a matrix where each of the columns is formed by the spectrum of a stellar template in the same format as $\boldsymbol{g}$, and $\boldsymbol{w}$ is a column vector with the number of stars that we have for each of the template stars. We do not account for dust obscuration nor emission lines.

The stars that are present in an SSP are defined by an isochrone as a function of its age and metallicity. When we combine a stellar library with an interpolator that allows us to interpolate between the spectra of the stars in the library (as a function of effective temperature, surface gravity and metallicity), we can create a spectrum for each of the isochrone stars. For an SSP, these are the spectra that form the columns of matrix $\mathbf{S}$. We normalize the spectra in matrix $\mathbf{S}$ to match the Johnson V-magnitude defined by the isochrone. Because an isochrone is a function of age t and

metallicity [M/H], the matrix is of the form $\mathbf{S}(t, [\mathrm{M/H}])$. The abundance pattern is currently not a free parameter in our model (see Conroy & van Dokkum 2012a, for a model where the abundance of various elements is allowed to vary).

The vector $\boldsymbol{w}$ (referred to as weights) are the unknowns in Equation 3.1. The IMF,

$$\xi(M) \equiv \frac{\mathrm{d}N}{\mathrm{d}M}, \tag{3.2}$$

and the weights are related through

$$\xi(m_j) = \frac{\boldsymbol{w}_j}{m_{\mathrm{high}} - m_{\mathrm{low}}}, \tag{3.3}$$

and vice versa through

$$\boldsymbol{w}_j = \int_{m_{\mathrm{low}}}^{m_{\mathrm{high}}} \xi(M)\mathrm{d}M. \tag{3.4}$$

In these equations, $\boldsymbol{w}_j$ is the number of stars for template j, m_j is the mass of that template star and m_{low} and m_{high} are the mass boundaries of this template.

Since Equation 3.1 is in general an ill-posed problem, solving it directly through a linear inversion may lead to very unrealistic and unphysical solutions. DTK16 therefore use a prior on the IMF to regulate the inversion of Equation 3.1. Using Equation 3.4, this IMF prior can be translated into a prior on the weights $\boldsymbol{w}_0$. Assuming Gaussian distribution functions of both the errors on the data and the variation of the weights around the mean of the prior, we can solve Equation 3.1 by minimizing

$$\begin{aligned} M(\boldsymbol{w}|\boldsymbol{g}, \mathbf{S}, \boldsymbol{w}_0, \mathbf{C}_{\mathrm{pr}}^{-1}) = &\frac{1}{2}(\mathbf{S}\,\boldsymbol{w} - \boldsymbol{g})^{\mathrm{T}}\mathbf{C}_{\mathrm{D}}^{-1}(\mathbf{S}\,\boldsymbol{w} - \boldsymbol{g}) \\ &+ \lambda \cdot \frac{1}{2}(\boldsymbol{w} - \boldsymbol{w}_0)^{\mathrm{T}}\mathbf{C}_{\mathrm{pr}}^{-1}(\boldsymbol{w} - \boldsymbol{w}_0), \end{aligned} \tag{3.5}$$

where $\mathbf{C}_{\mathrm{D}}^{-1}$ is the inverse noise covariance matrix and λ is the regularization parameter that balances the solution of $\boldsymbol{w}$ between that of the most likely solution and that of the prior solution. $\mathbf{C}_{\mathrm{pr}}^{-1}$ is the inverse covariance matrix describing the shape of the Gaussian distributed deviations of the weights from the prior on the weights $\boldsymbol{w}_0$ and is related to the regularization scheme

that is used. In this equation, the first term is related to the likelihood and the second (regularization) term puts a penalty on $\boldsymbol{w}$ for deviating from the prior on the weights $\boldsymbol{w}_0$. The balance between these two terms is set by the regularization parameter λ which is chosen such that it optimizes the posterior probability.

As shown by DTK16, the IMF prior $\boldsymbol{w}_0$ can be chosen as part of an IMF prior model family that is defined by a set of non-linear parameters. For example, we may assume that the IMF prior is a power law IMF which is defined by two non-linear parameters: its slope and its normalization. These non-linear IMF prior parameters can be sampled via Markov Chain Monte Carlo (MCMC) sampling techniques, and different choices of IMF prior parameters may be compared on the basis of the Bayesian evidence (MacKay 1992).

At the highest level of inference (second level of inference), we can marginalize over the weights and the non-linear parameters age, metallicity and IMF prior parameters to calculate the evidence for one particular model choice. This allows us to objectively compare different model ingredients, such as the choice of isochrones or the IMF prior model family, with each other on the basis of the evidence.

For more details with respect to the hierarchical Bayesian framework for SPS we refer the reader to DTK16.

3.2.2 New model features

In this section we describe the new features that we add to the model of DTK16. Among these features are a different method for enforcing positive weights, a multiplicative polynomial, an adaptive covariance matrix, the inclusion of multiple SSPs in our data model and including the velocity dispersion as a free parameter of our model. We also discuss a new sampling strategy where we first use a parameterized version of our code to find the most probable ages and metallicities of the SSPs included in the fit. Then we use the full version of the code to reconstruct the parameters related to the IMF.

3.2.2.1 Enforcing positive weights

A negative number of stars is unphysical, and therefore we force the model to consider solutions with positive weights only. DTK16 enforce positive weights by using a non-negative least squares (NNLS) method. We have

found that using NNLS can sometimes result in unrealistic solutions. We therefore choose to use a different approach for enforcing positive weights.

If the regularization parameter in our model is increased, we force the solution to become more and more equal to the prior on the weights (which is always positive). Therefore, above some threshold value of the regularization parameter the weights will always be positive. Here, we use this property of the regularization parameter and enforce positive weights by using an additional prior on the regularization parameter λ. This prior has a probability of one if all weights are positive and zero probability if any of the weights is below zero. With the additional prior on the regularization parameter we ensure a positive solution that is relatively close to the prior IMF while the model still has the freedom to deviate from the IMF prior if required by the data.

3.2.2.2 Multiplicative polynomial

Extinction, flux-calibration issues and systematic uncertainties in the model can potentially introduce differences between the continuum of the input spectrum and the continuum of the model spectrum. A possible solution to absorb these differences is to include a multiplicative polynomial in the fit. The relevance of including a multiplicative polynomial in the fit is discussed in e.g. Koleva et al. (2008).

Within the context of the hierarchical Bayesian framework discussed in Section 3.2.1, the inclusion of a multiplicative polynomial is complicated by the fact that the linear weights and the coefficients of the polynomial are degenerate. The weights and polynomial coefficients can therefore not be solved independently. To resolve this problem, we use an iterative procedure.

First, we determine the most probable weights without a multiplicative polynomial. These most probable weights are used to create a reconstructed spectrum and the fractional residual between the reconstructed spectrum and the input spectrum is determined. Then, we fit a multiplicative polynomial to the fractional residual. We apply this polynomial to the spectra in matrix $\mathbf{S}$ and determine the most probable weights again, now including the effect of a multiplicative polynomial.

To fit the residuals between the reconstructed spectrum and the input spectrum we use a tenth order (Legendre) polynomial $P_{10}(\lambda)$. According to Koleva et al. (2009), the optimal order of the polynomial depends on

wavelength range and resolution but in most cases $n = 10$ should be sufficient to get an unbiased determination of the parameters of a non-linear component, such as for example a stellar atmosphere or an SSP. In addition, Koleva et al. (2009) did not find a degeneracy between their multiplicative polynomials and the parameters of their non-linear components. We have also tested this for a number of SSPs that we distorted by a random polynomial. The results of this test suggest that the inference of age, metallicity and IMF slope of these SSPs is not affected by this distortion.

3.2.2.3 Age-metallicity reconstruction

To determine the most probable age and metallicity of an SSP, DTK16 consider a grid of ages and metallicities. For each of the points in this grid, the Bayesian evidence is determined by marginalizing over the IMF prior parameters. Under the assumption of a clear peak in the evidence of this age-metallicity grid and only minor degeneracy between the age, metallicity and IMF over the bulk of the evidence in age-metallicity space, the posterior of the weights and IMF prior parameters for the most probable age-metallicity grid point is used as a good approximation for the posterior of the weights and IMF prior parameters marginalized over all ages and metallicities.

Here we use a more general approach where we sample the age and metallicity of an SSP together with the IMF prior parameters. In principle, for every age-metallicity combination we have to create a set of stellar templates. However, since this process is computationally expensive, we use a discrete grid of ages and metallicities for which we a priori create a corresponding set of stellar templates. Within the model, any given combination of age and metallicity is then mapped to the closest discrete age-metallicity grid point and the corresponding set of stellar templates is used for the reconstruction. So although age and metallicity are still internally handled as discrete parameters within the model, this allows us to sample them as continuous parameters. The spacings of the age-metallicity grid points determine our resolution along these dimensions.

3.2.2.4 Adaptive covariance matrix

The standard noise covariance matrix in our model only takes into account the observational error on the input spectrum. In reality, there are additional uncertainties such as systematics in the model and uncertainties

related to the input spectrum being a CSP instead of an SSP. Therefore, we will in general underestimate the error budget of the model parameters. To model these additional uncertainties, we introduce an additional variance with value b. This parameter represents additional uniform variance that we do not take into account. If the original (diagonal) covariance matrix is given by $\mathbf{C}_{\mathrm{D,old}}$, the new covariance matrix becomes

$$\mathbf{C}_{\mathrm{D,new}} = \mathbf{C}_{\mathrm{D,old}} + b\mathbf{I}. \tag{3.6}$$

Note that this is still a diagonal covariance matrix. When we apply our model to real data, a more general covariance matrix, such as for example discussed by Czekala et al. (2015), might be required. However, since we need to calculate the inverse of the covariance matrix, which can be rather time consuming, we do not currently implement such a general covariance matrix.

3.2.2.5 Multiple SSPs

Real galaxies are CSPs. If we model a galaxy as if it is an SSP, this is an approximation that can introduce additional uncertainties. Depending on the star formation history (SFH) of the galaxy that we consider, this may or may not be a good approximation. The ETGs that we are interested in are in general expected to be characterized by an SFH consisting of an initial peak (old stellar population) followed by a small amount of more recent star formation (as observed in present-day ETGs by, e.g. Kaviraj et al. 2008; Monachesi et al. 2012; McDermid et al. 2015). Although the initial peak is probably modelled reasonably well by an SSP, the SFH is in reality extended. This may introduce a bias on the inferred age and metallicity of the SSP and potentially affect our inference of the IMF. To absorb some of these effects, we extend our model in such a way that it allows us to model a CSP as a combination of a given number of SSPs with varying age and metallicity. This approach allows us in principle to recover the age-metallicity distribution function that has the highest overall Bayesian evidence. It also allows for objective comparison between SSP models and CSP models with a varying number of SSP components (see also Section 3.5.1 where we test this for a number of mock CSP spectra).

Compared to our original approach, where the columns of matrix $\mathbf{S}$ correspond to the spectra of an SSP, the columns of matrix $\mathbf{S}$ now correspond to the spectra of multiple SSPs. For example, for N SSPs

matrix $\mathbf{S}$ becomes

$$\mathbf{S} = \begin{pmatrix} \mathrm{SSP}_1 & \cdots & \mathrm{SSP}_i & \cdots & \mathrm{SSP}_N \end{pmatrix}. \tag{3.7}$$

Similarly, the weights in Equation 3.1 become

$$\boldsymbol{w} = \begin{pmatrix} \boldsymbol{w}_{\mathrm{SSP}_1} \\ \vdots \\ \boldsymbol{w}_{\mathrm{SSP}_i} \\ \vdots \\ \boldsymbol{w}_{\mathrm{SSP}_N} \end{pmatrix}, \tag{3.8}$$

and the prior on the weights becomes

$$\boldsymbol{w}_0 = \begin{pmatrix} \boldsymbol{w}_{0,\mathrm{SSP}_1} \\ \vdots \\ \boldsymbol{w}_{0,\mathrm{SSP}_i} \\ \vdots \\ \boldsymbol{w}_{0,\mathrm{SSP}_N} \end{pmatrix}. \tag{3.9}$$

In principal, for every SSP that we take into account we can have a different IMF prior $\boldsymbol{w}_{0,\mathrm{SSP}_i}$. Here, we assume that the shape of the IMF prior is the same for all SSPs that we include in our model. However, the normalization of the IMF prior is a free parameter for every SSP because every SSP will contribute differently to the integrated light of the CSP. Note that although the shape of the IMF prior is assumed to be the same for all SSPs, the actual inferred IMF shape can change for different SSPs since the most probable weights are allowed to deviate from the prior for each SSP.

Altogether, every SSP introduces three additional non-linear parameters that have to be sampled by the MCMC sampler: its age, its metallicity and the normalization of the IMF prior, besides the IMF shape parameters which are assumed to be the same for all populations (hence no additional free parameters are introduced as compared to an SSP). The computational time scales are related to the increase in the number of parameters that have to be sampled and the increase of matrix $\mathbf{S}$ as we add more SSPs, which limits this approach to a few SSPs.

3.2.2.6 Velocity dispersion

In DTK16, we smooth our (mock) spectra to a given velocity dispersion. The stellar templates are then smoothed to the same velocity dispersion. Here, we include the velocity dispersion σ as an additional free parameter. We assume a Gaussian smoothing kernel in $\log \lambda$.

3.2.2.7 Parameterized version of model

In the original version of the model, we use the prior on the weights $\boldsymbol{w}_0$ to regularize the linear inversion of the most probable weights $\boldsymbol{w}_{\rm MP}$. However, for a given set of stellar templates $\mathbf{S}$ we can also use $\boldsymbol{w}_0$ to directly compute a corresponding prior spectrum $\boldsymbol{g}_0$ through

$$\boldsymbol{g}_0 = \mathbf{S}\,\boldsymbol{w}_0. \tag{3.10}$$

This prior spectrum can be compared to the data $\boldsymbol{g}$ on the basis of the likelihood

$$\mathcal{L}(\boldsymbol{g}_0|\boldsymbol{g},\mathbf{S}) = \frac{\exp[-E_{\scriptscriptstyle D}(\boldsymbol{g}_0|\boldsymbol{g},\mathbf{S})]}{Z_{\scriptscriptstyle D}}, \tag{3.11}$$

in which $Z_{\scriptscriptstyle D}$ is the normalization of the likelihood and

$$E_{\scriptscriptstyle D}(\boldsymbol{g}_0|\boldsymbol{g},\mathbf{S}) = \frac{1}{2}(\boldsymbol{g}_0-\boldsymbol{g})^{\rm T}\mathbf{C}_{\rm D}^{-1}(\boldsymbol{g}_0-\boldsymbol{g}). \tag{3.12}$$

We have developed a version of the code where we compare the input spectrum $\boldsymbol{g}$ directly to the prior spectrum $\boldsymbol{g}_0$ on the basis of the likelihood in Equation 3.11. We will refer to this version of the code as the parameterized version. The advantage of the parameterized version of the the code is that it is much faster then the original version. In the original version of the code, every iteration requires an iteration to find the most probable regularization parameter, and, for every step in this iteration, we have to perform a linear inversion to determine the most probable weights and the evidence. This process is time consuming. In the parameterized version of the code we only evaluate Equations 3.10 and 3.11, which is much faster but comes at the cost of less flexibility.

Similar to the full version of the code, we include a multiplicative tenth-order (Legendre) polynomial in the fit. In Section 3.2.2.8 we explain when we use the parameterized version of the code and when we use the full version of the code in our sampling strategy.

3.2.2.8 Sampling strategy

In the CSP version of the code we change our sampling strategy. First, we use the parameterized version of our code discussed in Section 3.2.2.7 to determine the most probable ages and metallicities of the SSPs that we include in our fit. As discussed in Section 3.2.2.3, the ages and metallicities of the SSPs are sampled together with the IMF prior parameters. Then, we fix the ages and metallicities of the SSPs to the most probable values from the first step and we sample the IMF prior parameters with the full version of our model (including linear inversion). This initial sampling procedure is much faster than when we sample the complete space of ages, metallicities and IMF prior parameters with the full version of our model but assumes that there is little covariance between the age-metallicity parameters and small deviations from the IMF prior. We also use the parameterized version of our code to determine the velocity dispersion of a spectrum.

As in DTK16, we have implemented our code as a pipeline of `cosmoSIS` (Zuntz et al. 2015). We use `Multinest` (Feroz & Hobson 2008; Feroz et al. 2009, 2013) to calculate evidences and obtain a posterior sample of the sampled parameters. If necessary, the posterior sampling is refined with `emcee` (Foreman-Mackey et al. 2013).

3.3 Stellar templates

In this section we describe how we construct the stellar templates that form the columns of the stellar-template matrix $\mathbf{S}$. We also describe a binning procedure where we combine multiple stellar templates into one new template. This binning procedure allows us to reduce the sizes of the stellar templates matrices and to speed up the code.

3.3.1 Isochrones

The stellar templates that we use in this work are created on the basis of the stellar parameters provided by the Parsec isochrones (Bressan et al. 2012) extended with the thermally pulsating - asymptotic giant branch (TP-AGB) tracks from Marigo et al. (2013) and Rosenfield et al. (2016). This is the most recent version of the Parsec isochrones and includes several updates to the input physics of the Padova stellar evolution code, as well as an improved treatment of low mass dwarfs as described in Chen et al. (2014).

Each isochrone provides us with the initial masses, effective temperatures, surface gravities and Johnson V-band magnitudes of the stars that are present in an SSP.

3.3.2 Stellar library and interpolator

To transform the stellar parameters of an isochrone into a set of spectra, we combine a stellar library with an interpolator. We use the empirical MILES stellar library (Sánchez-Blázquez et al. 2006) which consists of approximately 1000 stars in the wavelength range 3540.5 - 7409.6 Å. To interpolate between the stars in this library, we use the same interpolator as in DTK16. This is a local interpolator based on Vazdekis et al. (2003). Since we now include a multiplicative polynomial in our model, we do not apply the polynomial correction of the MILES stars discussed in DTK16 but use the spectra of the MILES library directly.

The interpolator $S_\lambda(T_{\rm eff}, \log g, [{\rm Fe/H}])$ allows us to create a stellar spectrum for any given set of stellar parameters $T_{\rm eff}$, $\log g$ and [Fe/H]. With this interpolator we create a stellar-template matrix $\mathbf{S}$ for each of the grid points in our age-metallicity grid. Since the isochrones are defined on the basis of $[{\rm M/H}] = \log(Z/Z_\odot)$ but the MILES stars only have measured values of [Fe/H], we assume that [M/H] = [Fe/H].

3.3.3 Isochrone binning

Since the parameters of one isochrone star to the next isochrone star do in general not change drastically, the interpolated spectra are expected to vary smoothly as a function of stellar mass. The number of stars in a Parsec isochrone is typically on the order of a few hundred and can be as high as a thousand. Having such a large number of templates makes the matrices in our model unnecessarily large which in turn makes the code unnecessary slow. We therefore reduce the number of stellar templates in the matrices by binning them together.

The sampling of the isochrone stars increases as a function of evolutionary phase and is particularly high for the TP-AGB. Hence, we let the number of stars that we bin together depend on the evolutionary phase. For the main sequence we combine every two spectra into a new template, for the TP-AGB we combine every eight spectra into a new template and for the other evolutionary phases we combine every three spectra into a new template. With these numbers, we reduce the computational time scales

significantly and are still able to model a spectrum with up to $N = 6$ SSPs (see Section 3.5.1) in a reasonable amount of time[2].

When we bin the spectra, we assume a Salpeter IMF. In Appendix 3.A we show that this assumption has only a very small effect on our SSP spectra and that it does not affect our inference of the IMF.

3.4 Composite stellar population mock spectra

In this section we describe the CSP mock spectra that we use to test our model. First, we discuss the semi-analytic model from which we extract the SFHs of our CSP mock spectra. Then we describe how we use these SFHs to construct a set of CSP mock spectra with a variable IMF.

3.4.1 SFHs from semi-analytic models

Semi-analytic models (SAMs) are an ideal tool to create realistic SFHs of galaxies. The main advantage of SFHs from SAMs over more idealized SFHs (e.g., delayed τ models) is that they can include increasing and decreasing phases, as well as bursts of star formation. Within SAMs, simplified analytical but physically motivated recipes are used to track dynamical and astrophysical processes. This happens on entire galaxy scales, rather than kpc or even pc scales (the physical scales below which cosmological and zoom-in hydro-simulations apply subgrid physics). This ensures an inexpensive runtime for SAMs, which makes them a powerful tool to generate statistically significant model galaxy populations covering a wide range in galaxy properties (for a recent review see Somerville & Davé 2015).

In this work we focus on the SAM first presented in Somerville & Primack (1999) and further updated by Somerville et al. (2008), Somerville et al. (2012), and Porter et al. (2014b). This SAM includes recipes to track the hierarchical clustering of dark-matter haloes, shock heating and radiative cooling of gas, supernova feedback, star formation, black-hole growth and active galactic nuclei feedback, the enrichment of the interstellar and intracluster medium with metals, mergers of galaxies, starbursts, and the evolution of stellar populations. Porter et al. (2014b) included recipes to track the effect of disk-instabilities and mergers on the structural evolution

[2]Typically, modelling a spectrum with $N = 6$ SSPs takes on the order of 3-4 hours using 64 cores.

of galaxies. Of particular importance within the context of this work, Porter et al. (2014b) included recipes to track the line-of-sight velocity dispersion of the stars in modelled galaxies. These models have proven to be fairly successful in reproducing local and high-redshift statistical galaxy properties such as stellar mass functions, gas masses, dust masses, sizes, SFRs, and also quenched fractions and structural properties such as Sérsic index(Somerville et al. 2008, 2012, 2015; Porter et al. 2012; Popping et al. 2014, 2016a,b; Lang et al. 2014; Brennan et al. 2015, 2017). We refer the reader to the aforementioned works for more detailed descriptions of the utilized model.

We make use of merger trees extracted from the Bolshoi N-body dark matter simulation (Klypin et al. 2011; Trujillo-Gomez et al. 2011; Rodríguez-Puebla et al. 2016). Dark matter haloes were identified using the `ROCKSTAR` algorithm (Behroozi et al. 2013). This simulation is complete down to haloes with a virial velocity of $V_{\rm circ} = 50\,{\rm km\,s^{-1}}$, with a force resolution of $1\ h^{-1}$ kpc and a mass resolution of $1.9 \times 10^8\,{\rm M}_\odot$ per particle. The results presented assume a ΛCDM cosmology with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $h = 0.7$, $\sigma_8 = 0.82$, $n = 0.95$ and a cosmic baryon fraction of $f_b = 0.1658$. These parameters were chosen to match those adopted in the Bolshoi simulation (Klypin et al. 2011) and are consistent with the *Wilkinson Microwave Anisotropy Probe* (WMAP) results (Komatsu et al. 2011).

The semi-analytic model assumes a Chabrier (2003) IMF. Note that this IMF is only assumed within the SAM. As described in Section 3.4.2, the spectra of the mock CSPs are determined under the assumption of variable IMF. Therefore, the IMF assumed in the SAM and the IMF used to determine the spectra are not consistent with each other. In principal, the assumption of a universal Chabrier (2003) IMF within the SAM might affect the enrichment of the ISM with metals, the mass-loading factor and the heating rate from massive stars (supernovae and winds), and eventually also the stellar mass assembly of a galaxy. However, within the context of this work we only want to construct a complex age-metallicity distribution and not necessarily a precise one and therefore these effects are not expected to affect our conclusions.

Galaxies drawn from the SAM are selected to be passive, with specific star formation rates ${\rm SFR}/M_* < 10^{-11}\,{\rm yr^{-1}}$, to match approximately the selection criteria of Spiniello et al. (2014). We then average the SFHs of these passive galaxies in $40\,{\rm km\,s^{-1}}$-wide bins centred at velocity dispersions

of 150, 190, 230, 270, and 310 $\mathrm{km\,s^{-1}}$. The resulting SFHs suggest that some of the stars that have formed are relatively young ($t < 8$ Gyr) with a very low metallicity ([Fe/H] $= -2.4$). We suspect that this is connected to a mode of star formation which is related to infalling gas from the intergalactic medium or it could be accretion from a poorly resolved, low-mass satellite. Since this kind of stellar population is not seen in present-day galaxies, we manually set the stellar mass in bins with $t < 8$ Gyr and [Fe/H] $= -2.4$ to zero.

Trager & Somerville (2009) compare SSP-equivalent ages between a mock catalogue of SAM galaxies and a number of observational samples and show that these ages approximately agree. Although the observational samples do not allow the reconstruction of the full SFH, this would imply that the amount of recent star formation produced by the SAM is approximately correct. Fontanot et al. (2009) compare the mean mass-weighted stellar age as a function of stellar mass of three SAMs with the observational results of Gallazzi et al. (2005) and show that the agreement for massive galaxies is very good. Porter et al. (2014a) compare mass weighted ages of ETGs in their SAM with the luminosity-weighted ages of the sample of SDSS ETGs by Graves et al. (2009) as a function of effective radius and velocity dispersion. The SDSS galaxies are found to be younger than the SAM galaxies, but this difference most likely originates from comparing luminosity-weighted with mass-weighted ages, where the former is known to be biased towards younger ages, i.e. dominated by young stars (Trager & Somerville 2009). However, the trends recovered between velocity dispersion, age and metallicity are very similar for both samples. These results suggest that the SFHs that we use in this work may not be completely accurate but are at least a realistic representation of what might be expected for ETGs.

3.4.2 CSP mock spectra with a variable IMF

The SAMs provide us with an average SFH for galaxies in five different bins of velocity dispersion (150, 190, 230, 270 and 310$\mathrm{km\,s^{-1}}$). The average SFH for the different velocity dispersion bins is visualized in Figure 3.1. For each of the velocity dispersion bins, the average SFH is given as the total current mass of stars in a predefined age-metallicity grid. This age-metallicity grid has ages in the range $t = \{0.1, 13.5\}$ Gyr and metallicities in the range [M/H] $= \log(Z/Z_\odot) =$ [Fe/H] $= \{-2.4, 0.47\}$ with $Z_\odot = 0.0152$ (Bressan

Table 3.1 – Summary of the five different CSP mock spectra that we consider. Each of the mock spectra corresponds to a different velocity dispersion bin, the SFH is derived from the SAM and the IMF is assumed to be a single power law IMF. The first column gives the name of the CSP mock spectrum, the second column the velocity dispersion bin and the last the column the IMF slope. For every spectrum there are three versions corresponding to SNRs of 75, 150 and 300, respectively.

name	σ [km s^{-1}]	α
`CSP150`	150.0	1.84
`CSP190`	190.0	2.08
`CSP230`	230.0	2.27
`CSP270`	270.0	2.43
`CSP310`	310.0	2.57

et al. 2012). The spacings of this grid are shown in the last panel of Figure 3.1. Note that the age-metallicity grid of the isochrones is chosen to match this grid.

We assume that the IMF of our CSP mock spectra is a single power law IMF. Furthermore, we assume that the slope α of this IMF varies as a function of velocity dispersion according to the empirical IMF slope–velocity dispersion relation of Spiniello et al. (2014):

$$\alpha = 2.3 \log \sigma_{200} + 2.13, \tag{3.13}$$

where σ_{200} is the velocity dispersion in units of 200 km s^{-1}.

For every bin in velocity dispersion we create a corresponding CSP mock spectrum. To create these CSP mock spectra, we first create an SSP spectrum SSP_{ij} for each of the age-metallicity grid points. These SSP spectra are calculated using Equation 3.1, where the stellar templates $\mathbf{S} = \mathbf{S}(t_i, [\mathrm{Fe/H}]_j)$ are those of an SSP and the weights $\boldsymbol{w}$ are determined through Equation 3.4 assuming a single power law IMF $\xi(\sigma)$ where the slope is given by Equation 3.13. The SSP spectra are normalized to a total (current) stellar mass of 1 $\mathrm{M}_\odot$.

The CSP mock spectrum $\boldsymbol{s}_{\mathrm{CSP}}$ is then determined as the weighted sum over all SSP spectra:

$$\boldsymbol{s}_{\mathrm{CSP}} = \sum_{i,j} c_{ij}\, \mathrm{SSP}_{ij}, \tag{3.14}$$

where the weights c_{ij} of each of the SSP spectra follow from the SFHs of the SAM. We smooth the CSP mock spectra according to the corresponding

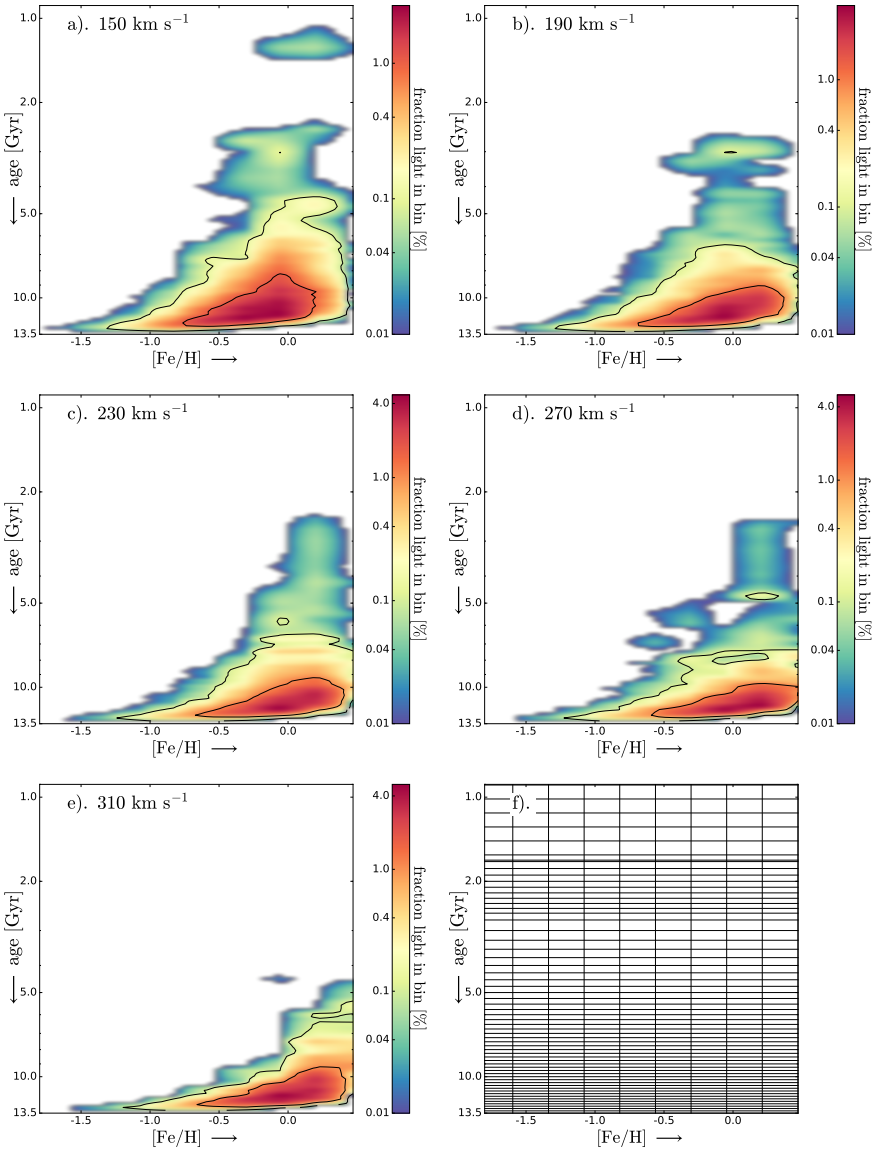

Figure 3.1 – Average SFHs of galaxies in the SAM binned by velocity dispersion. For each point in the age-metallicity grid, we determine the fractional contribution of the integrated SSP spectrum to the integrated CSP spectrum and the (colour) contours in this plot are interpolated on the basis of these values. The SFHs in panel a-e correspond to different velocity dispersions. Panel (f). represents the adopted age-metallicity grid in the same range as the previous panels.

model velocity dispersion. This gives us a total of five mock CSP spectra, which are summarized in Table 3.1. As a final step, we add Gaussian noise to the spectra to mimic observations with signal-to-noise ratios (SNRs) of 75, 150 and 300 per bin, so that we have three version of every CSP mock spectrum corresponding to different SNRs.

3.5 Results

In this section we present the analysis of the CSP mock spectra with our model. We focus first on `CSP150` and `CSP310`. These mock spectra are characterized by, respectively, the most and least extended SFH, as can be seen in Figure 3.1. Then we extend our analysis to the other mock spectra and show that we can reconstruct the underlying variable IMF. We use the sampling strategy discussed in Section 3.2.2.8. First we use the parameterized version of our model to determine the most probable ages and metallicities of the SSPs and the velocity dispersion of the spectrum. Note that the velocity dispersion is derived by broadening the stellar templates in matrix $\mathbf{S}$ with σ as a free parameter. Then we use the full version of the model to sample the non-linear IMF prior parameters and the additional covariance parameterized by b as discussed in Section 3.2.2.4.

For all of the analyses that we present here, we assume a single power law IMF prior. The regularization scheme that we use penalizes the relative deviation of the weights $\boldsymbol{w}$ from the prior on the weights $\boldsymbol{w}_0$ and is specified by

$$\mathbf{C}^{-1}_{\mathrm{pr},ii} = \frac{1}{{\boldsymbol{w}_{0,i}}^2} \tag{3.15}$$

where $\mathbf{C}^{-1}_{\mathrm{pr}}$ is a diagonal matrix.

3.5.1 The required number of SSPs

The Bayesian context of our model allows us to determine the most probable number of SSPs that is required to fit the data. As discussed in Section 3.2, our model is now able to fit a spectrum with a combination of multiple SSPs. For a given number of SSPs, if we marginalize over all the free parameters in the model we can determine the evidence for the given number of SSPs. Running the model multiple times allowing for different numbers of SSPs

therefore allows us to evaluate the Bayesian evidence as a function of the number of SSPs included in the fit. In principle, one can do this for an arbitrary number of SSPs. However, due to computational costs we limit ourselves to a maximum of $N = 6$ SSPs.

Figure 3.2 shows the difference in log evidence for `CSP150` and `CSP310` as a function of the number of SSPs and compared to the evidence determined with one SSP. The different lines in Figure 3.2 correspond to different SNRs, and one can clearly see that the most probable number of SSPs depends on the SNR of the spectrum. For SNR=75, the most probable number of SSPs is about 2-3, while for SNR=300 the evidence continues to rise to about 5-6 SSPs. This SNR dependence is not surprising, since spectra with higher SNR contain more information and are in general better able to discriminate between models.

In addition to the SNR dependence, Figure 3.2 also shows that the most probable number of SSPs for `CSP310` is lower than it is for `CSP150` for spectra with the same SNR. This difference is most likely explained by the more extended SFH of `CSP150`. In general, we expect that the more complicated the SFH the more SSPs are required to correctly fit the spectrum of a CSP. Another explanation might be the higher velocity dispersion of `CSP310`. A higher velocity dispersion implies that more information in the spectrum is lost due to the smoothing, and as a consequence it might become more difficult to distinguish different SSPs.

Figure 3.2 shows that we can gain significantly by going from one SSP to two SSPs. For all the CSP mock spectra that we consider, the difference in log evidence between two subsequent number of SSPs is the largest for going from one to two SSPs and ranges from a difference in log evidence of approximately 25 to approximately 700. According to Jeffreys (1961), a difference in log evidence of more than ten can be considered as strong evidence in favour of a particular model. Therefore, for all SSPs that we considered there is strong evidence to include two SSPs instead of one. Allowing for more than two SSPs may still improve the evidence, but the differences in log evidence are not as conclusive as going from one to two SSPs, though this might strongly depend on the SFH. The horizontal lines in Figure 3.2 indicate a difference of $\Delta \log \text{evidence} = -10$ with the peak value of the evidence. After the evidence curve crosses this horizontal line for the first time, there is no strong evidence to include any additional SSPs in the fit. Typically, for SNR = 75 there is strong evidence to include 2 SSPs, for SNR = 150 there is strong evidence to include 2-3 SSPs and for

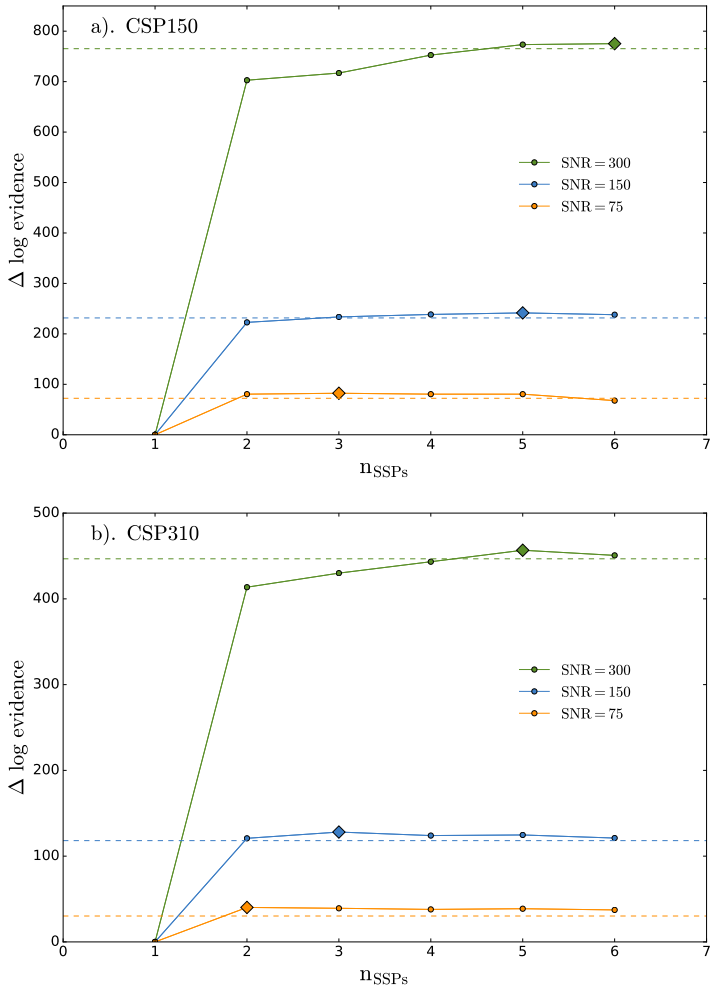

Figure 3.2 – Difference in log evidence as a function of the number of SSPs included in the analysis. The difference in log evidence is determined with respect to the evidence obtained with one SSP such that for $n_{\mathrm{SSPs}} = 1$, $\Delta \log \mathrm{evidence} = 0$. Panel (a) corresponds to `CSP150` and panel (b) to `CSP310`. Different lines correspond to different SNRs as indicated in the legend. The diamond data point represent the number of SSPs for which the evidence reaches a maximum. The horizontal lines correspond to a difference of $\Delta \log \mathrm{evidence} = -10$ with the peak value of the evidence, indicating for how many SSPs there is strong evidence to include them in the fit. For example, there is strong evidence for including five SSPs for `CSP150` at SNR=300 while at SNR=75 there is only strong evidence for including two SSPs.

SNR=300 there is strong evidence to include 5 SSPs. (See also Appendix 3.B for a test where we reconstruct the spectra of three mock SSPs with and without the original stellar templates to assess the level of the effect of systematic uncertainties on the difference in log evidence.)

In Figure 3.3 we show the reconstructed ages and metallicities of the SSPs used to fit `CSP150-150`[3] and where we allow for 1-6 SSPs. For the analysis with one SSP, the age of the SSP is clearly biased towards younger ages. Light-weighted SSP-equivalent ages are known to be biased towards younger ages and to underestimate the age of a CSP (Trager et al. 2000; Serra & Trager 2007; Trager & Somerville 2009), and we expect that the bias in Figure 3.3a is similar to this more generally known bias. However, for two or more SSPs the most prominent SSP (biggest dot) falls well within the central contour of the SFH of `CSP150`. Although the ages of the different SSPs in one fit can be quite similar, it is interesting to see that the model almost never selects two SSPs with the same metallicity. This may be related to the coarser sampling of the metallicity grid, and SSPs at different metallicities may also be used in part to absorb the effect of SSPs at different ages due to the age-metallicity degeneracy.

3.5.2 Reconstructed IMFs

As a next step we consider the reconstruction of the IMF for `CSP150-150`. To determine the reconstructed IMF we first sample the IMF prior parameters at fixed ages and metallicities of the SSPs[4]. Then we select the maximum-a-posteriori (MAP) set of IMF prior parameters to construct a prior on the weights and we use this prior to calculate the most probable weights. Finally, using Equation 3.3, we convert the most probable weights into a most probable reconstructed IMF. The error bars on the reconstructed IMF are derived on the basis of the posterior sample of the weights.

The reconstructed IMF of `CSP150-150` is shown in Figure 3.4 for models with $N = 1, 2, 3$ SSPs. For the analysis with one SSP, the most probable IMF deviates significantly from the prior IMF and the original IMF. This already shows that, with one SSP, a single power law IMF is not able to provide a fit to the data that is consistent with the specified noise level. If

[3]Here we use `CSP150-150` to refer to the version of `CSP150` with SNR = 150.

[4]The ages and metallicities of the SSPs are determined with the parameterized version of the model in a previous step.

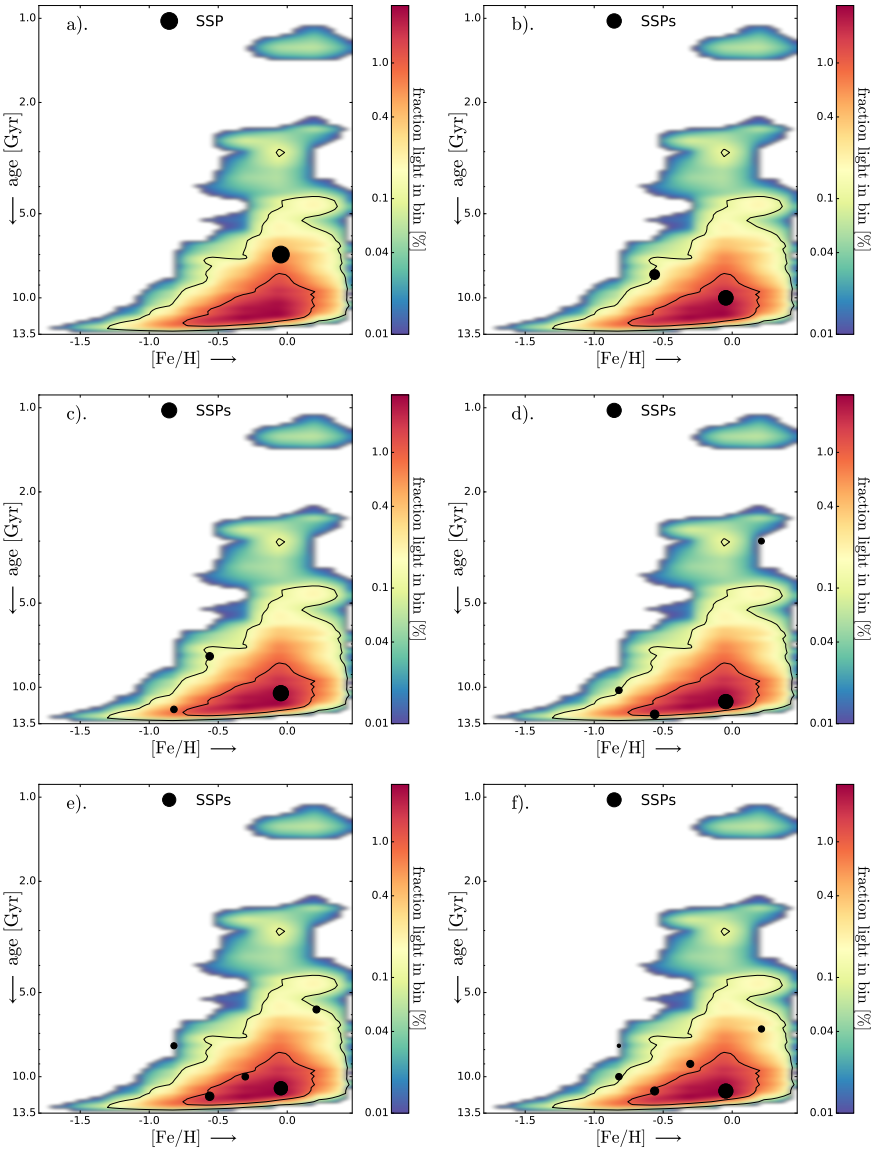

Figure 3.3 – Reconstructed ages and metallicities of the SSPs used to fit `CSP150-150`. The colour coding and contours correspond to the SFH of Figure 3.1a and the black dots represent the ages and metallicities of the SSPs fitted to the spectrum. The bigger the dot, the higher the relative contribution of that SSP to the integrated spectrum. Different panels correspond to a different number of allowed SSPs in the fit: (a). 1 SSP, (b). 2 SSPs, (c). 3 SSPs, (d). 4 SSPs, (e). 5 SSPs, and (f). 6 SSPs.

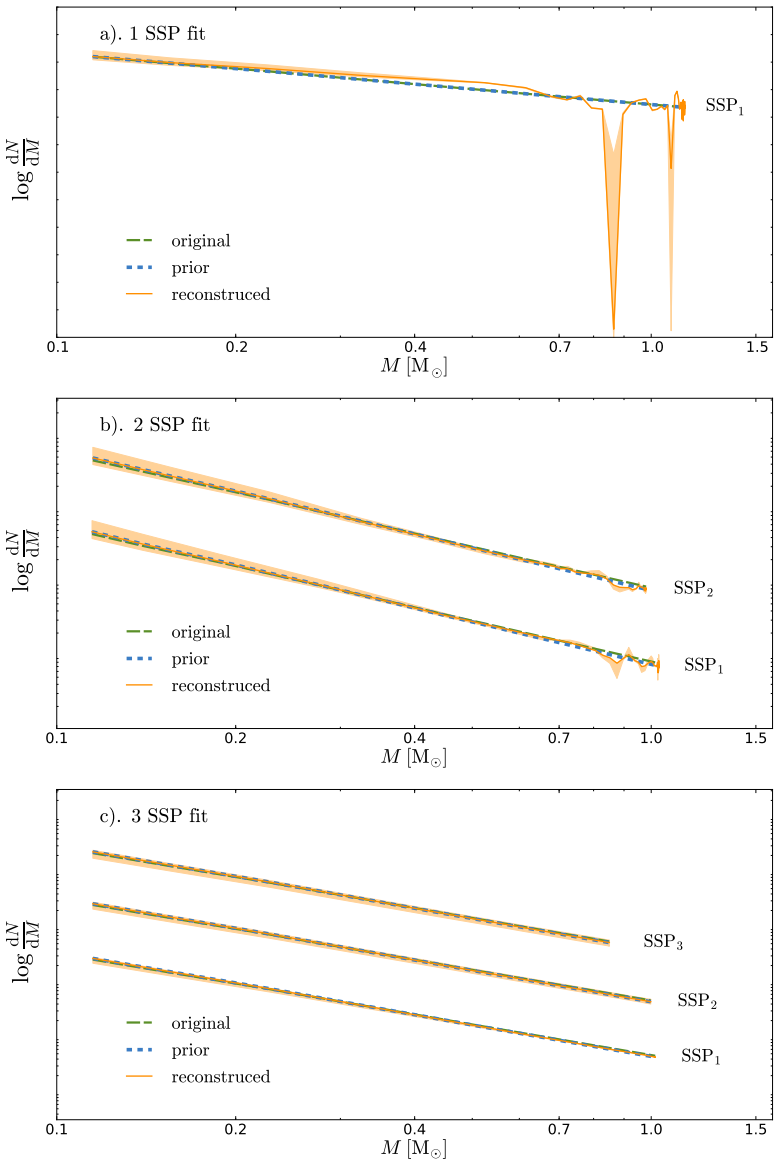

Figure 3.4 – Reconstructed IMF of `CSP150-150`. The *green-long-dashed* line represent the original IMF, the *blue-short-dashed* line the prior IMF and the *orange* line the most probable reconstructed IMF. The *shaded-orange* region represents the one sigma confidence interval of the reconstructed IMF and is derived from the distribution of most probable weights of the posterior sample. In panel (a) we allow for one SSP in the fit, in panel (b) for two SSPs and in panel (c) for three SSPs.

such a model would have been able to fit the data there would be no need for the model to deviate from the prior. Of course it makes sense that this model does not fit the data since our spectrum in reality represents a CSP with an extended SFH.

For two SSPs the reconstructed IMF is already much closer to the prior and the original IMF, although there are still some wiggles at the high-mass end. When we use three (or more) SSPs in the fit, the prior IMF, the original IMF and the reconstructed IMF become very similar. On the basis of this similarity we might conclude that three SSPs are sufficient to fit the spectrum of `CSP150-150`. According to Figure 3.2, for this particular case the evidence continues to increase a little bit more for $n_{\mathrm{SSPs}} > 3$ but the difference in log evidence between $N = 3$ and $N = 4$ SSPs is 4.9 which according to Jeffreys (1961) is not considered to be substantial evidence in favour of $N = 4$ SSPs. Three SSPs therefore seem to be sufficient here.

3.5.3 Reconstructed spectra

Finally, we consider the reconstructed spectra of `CSP150-150`. To determine the reconstructed spectrum we start with the MAP values of the sampled non-linear parameters and we use these values to calculate the most probable weights. Then we multiply the stellar-template matrix with the most probable weights to obtain the reconstructed MAP spectrum.

The reconstructed MAP spectrum of `CSP150-150` is shown in Figure 3.5. In the first panel, we use only one SSP for the fit whereas in the second panel we use five SSPs which is the maximum that we find in the evidence as a function of the number of SSPs. Looking at the reconstructed spectra for one and for five SSPs, at first glance there does not seem to be a difference. However, if one carefully looks at the residuals there appears to be a wave-pattern in the residuals when we use one SSP. If we use five SSPs, this pattern is far less prominent. To visualize this noise-pattern we also show a smoothed version of the residuals in Figure 3.5.

3.5.4 Reconstructing the variable IMF

As discussed in Section 3.4, the IMF of our CSP mock spectra varies as a function of velocity dispersion. All of our mock spectra have a single power law IMF for which the IMF slope is given by Equation 3.13. In this section, we try to reconstruct the underlying trend of the IMF from our mock spectra for different SNRs.

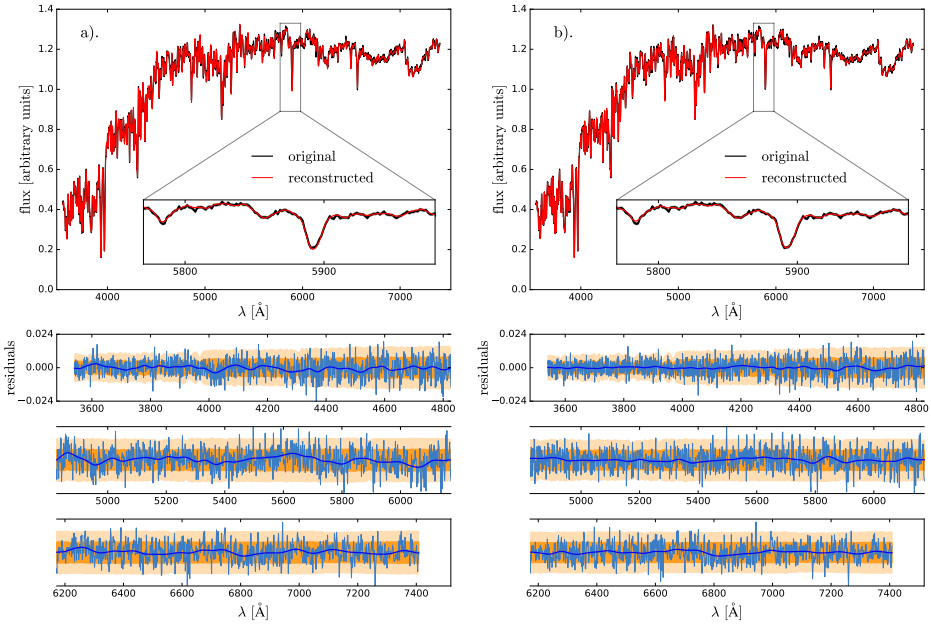

Figure 3.5 – Reconstructed spectrum of `CSP150-150`. The *black* line represents the original spectrum and the *red* line the reconstructed spectrum. In panel (a) we use one SSP for the fit and in panel (b) we use five SSPs. The bottom plots represent the residuals in *blue*. Also shown is a smoothed version of the residuals with the *thick-blue* lines. The *shaded-orange* region represent the one and two sigma errors of the spectrum and are derived from the covariance matrix $\mathbf{C}_{\mathrm{D}}$.

We analyse all of our mock spectra using the sampling strategy discussed in Section 3.2.2.8. First we derive the most probable ages and metallicities as well as the velocity dispersion with the parameterized version of the model. We then fix these parameters and run the full version of the model to calculate the evidence and obtain posterior samples of the IMF prior parameters, the weights and the covariance parameter b. For each of the mock spectra, we repeat the analysis with $N = 1 - 6$ SSPs, and we select the number of SSPs that maximizes the evidence.

The results of our analysis are given in Table 3.2. This table summarizes the number of SSPs that maximizes the evidence, the ages and metallicities of these SSPs, the reconstructed IMF slope, the reconstructed velocity dispersion and the reconstructed covariance parameter for each of our CSPs. The results in Table 3.2 confirm our results in Section 3.5.1: spectra with

Table 3.2 – Reconstructed parameters of CSP mock spectra for different SNRs. The different columns of this table give the name of the CSP, the number of SSPs that maximizes the evidence, the ages of these SSPs, the metallicities of these SSPs, the input value of the IMF slope that was used to create the spectrum, the reconstructed IMF slope, the reconstructed velocity dispersion, and the reconstructed covariance parameter. Errors on the IMF slope correspond to the 16th and 84th percentile of the posterior distribution, errors for σ and $b_{\rm cov}$ correspond to the standard deviation of the posterior sample. Note that with `CSP150-75` we refer to `CSP150` with SNR = 75.

name	$n_{\rm SSPs}$	SSP ages [Gyr]	SSP [Fe/H]s	$\alpha_{\rm in}$	$\alpha_{\rm rec}$	$\sigma_{\rm rec}$	$b_{\rm cov,rec}$
`CSP150-75`	3	10.5, 9.5, 9.0	-0.05, -0.56, -1.08	1.84	$1.76^{+0.12}_{-0.13}$	150.0 ± 1.1	0.02 ± 0.01
`CSP190-75`	3	9.8, 12.0, 11.0	-0.05, -0.56, -0.21	2.08	$2.08^{+0.11}_{-0.11}$	190.9 ± 1.2	0.02 ± 0.01
`CSP230-75`	4	11.5, 11.8, 8.3, 10.8	-0.05, -0.21, -0.30, -1.08	2.27	$2.28^{+0.10}_{-0.10}$	233.3 ± 1.4	0.02 ± 0.01
`CSP270-75`	3	12.3, 11.0, 7.5	-0.05, 0.21, -0.56	2.43	$2.39^{+0.13}_{-0.11}$	269.8 ± 1.5	0.01 ± 0.01
`CSP310-75`	2	9.8, 12.0	0.21, -0.30	2.57	$2.50^{+0.10}_{-0.12}$	312.6 ± 2.1	0.01 ± 0.01
`CSP150-150`	5	11.0, 10.0, 5.8, 11.8, 7.8	-0.05, -0.30, 0.21, -0.56, -0.82	1.84	$1.88^{+0.07}_{-0.07}$	150.0 ± 0.5	0.01 ± 0.01
`CSP190-150`	4	13.0, 8.0, 10.0, 13.3	-0.05, 0.21, -0.56, -0.82	2.08	$2.06^{+0.04}_{-0.05}$	188.8 ± 0.6	0.01 ± 0.01
`CSP230-150`	5	11.0, 11.5, 12.0, 4.5, 9.8	-0.05, 0.21, -0.56, 0.21, -1.08	2.27	$2.26^{+0.06}_{-0.06}$	229.8 ± 0.7	0.02 ± 0.01
`CSP270-150`	4	10.5, 10.8, 11.0, 12.0	0.21, -0.05, -0.56, 0.47	2.43	$2.43^{+0.04}_{-0.05}$	268.7 ± 0.8	0.01 ± 0.01
`CSP310-150`	3	10.0, 12.3, 11.5	0.21, -0.05, -0.56	2.57	$2.59^{+0.04}_{-0.04}$	310.1 ± 1.0	0.02 ± 0.01
`CSP150-300`	6	11.5, 8.3, 9.8, 10.8 9.5, 4.0	-0.30, -0.05, 0.21, -0.56 -1.08, -1.08	1.84	$1.80^{+0.03}_{-0.03}$	150.5 ± 0.3	0.02 ± 0.02
`CSP190-300`	5	10.5, 10.3, 10.0, 12.0, 11.5	-0.05, 0.21, -0.30, -0.56, -1.6	2.08	$2.12^{+0.02}_{-0.02}$	190.2 ± 0.3	0.02 ± 0.01
`CSP230-300`	4	10.0, 11.0, 11.8, 7.5	0.21, -0.05, -0.30, -0.56	2.27	$2.32^{+0.02}_{-0.03}$	229.7 ± 0.4	0.02 ± 0.01
`CSP270-300`	5	10.5, 12.3, 11.8, 6.8, 1.8	-0.05, 0.21, -0.56, 0.47, -0.05	2.43	$2.46^{+0.03}_{-0.02}$	270.5 ± 0.4	0.01 ± 0.01
`CSP310-300`	5	11.3, 11.5, 8.0, 13.0, 12.3	-0.05, 0.21, 0.46, -0.30, -0.82	2.57	$2.60^{+0.03}_{-0.02}$	310.3 ± 0.5	0.02 ± 0.02

higher SNRs and spectra with more extended SFHs in general require more SSPs. Since the SFHs of the CSPs are extended, we cannot directly compare the reconstructed ages and metallicities of the fitted SSPs. However, if we consider the ages and metallicities of the SSPs in Figure 3.1, they seem to cover the distribution function nicely. We want to emphasize that the aim of including multiple SSPs in the fit is not to reconstruct the SFH perfectly but to ensure that the determination of the IMF is not biased as a consequence of assuming only one SSP. In that respect, the age-metallicity distribution can be regarded for the purpose of this paper as nuisance parameters that are marginalized over. However, our analysis shows that even high-SNR spectra typically contain information of at most half a dozen distinct stellar-population regions in their space of ages and metallicities. This makes full-Bayesian modelling of these spectra to infer their IMFs and SFHs a tractable problem on present-day computers.

The reconstructed velocity dispersions in Table 3.2 are consistent with their input values. We have parameterized the extra covariance term by $b_{\rm cov}$, which is the extra (constant) covariance in terms of the median of the original covariance matrix, i.e.

$$b = b_{\rm cov} \cdot \mathrm{median}\left(\mathbf{C}_{\rm D}\right). \tag{3.16}$$

From the results in Table 3.2 we can see that $b_{\rm cov}$ is at most 0.02 so that the additional covariance is in general not more than 2% of the median of the original covariance matrix. The additional covariance term does therefore not seem to play an important role here, but it might become more important if we deal with real data where we have more systematic uncertainties.

The reconstructed IMF slopes are consistent with the input values. In Figure 3.6 we show the reconstructed IMF slopes (orange data points) together with the IMF slope–velocity dispersion relation specified in Equation 3.13 for SNRs of 75, 150 and 300, respectively. These IMF slopes are derived by using the number of SSPs that maximizes the evidence. The results show that for all SNRs we are able to correctly reconstruct the IMF slope–velocity dispersion relation. As expected, for lower SNRs the scatter in this relation increases. For SNR=300, the error bars seem to be slightly underestimated, or the results are slightly biased, but the overall trend is correct and the blue line in Figure 3.6c would still fall within the two sigma error bars of the data points.

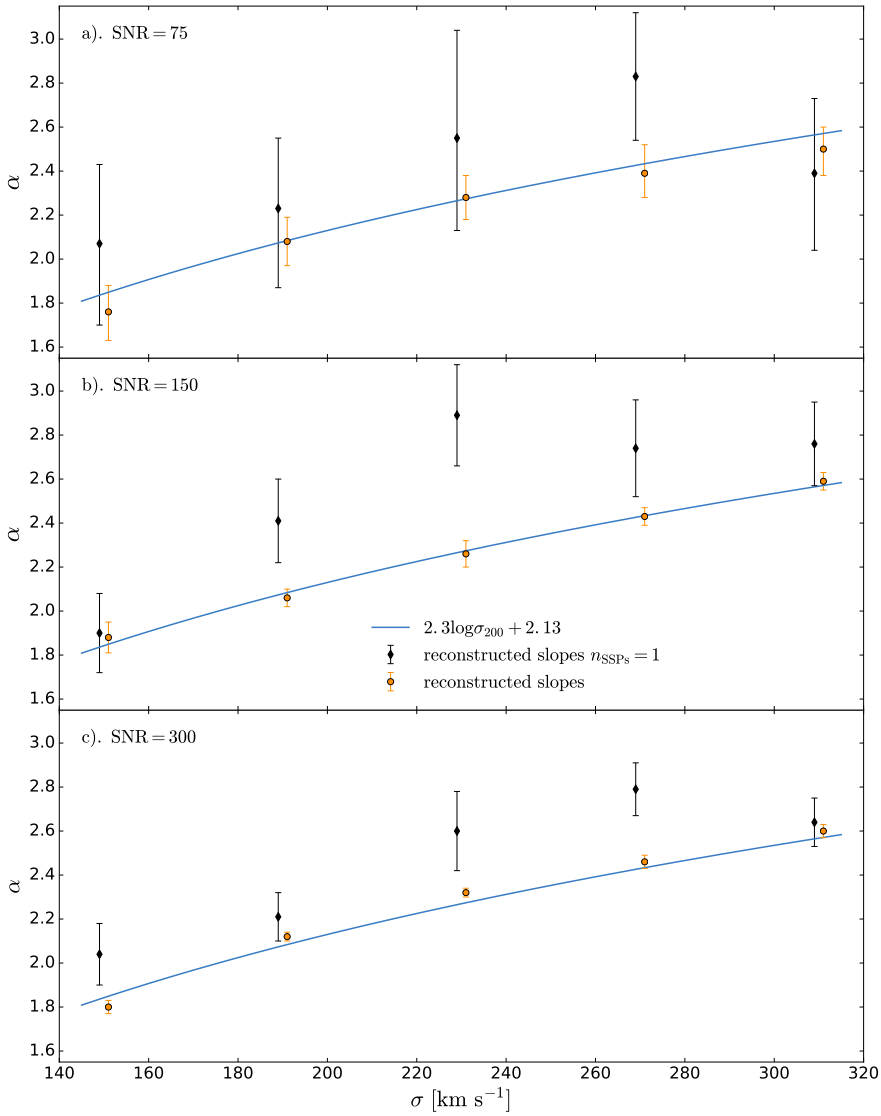

Figure 3.6 – Reconstructed IMF slopes versus (true) velocity dispersion for CSP mock spectra with different SNRs. Data points represented by *black diamonds* are derived by using one SSP whereas for the *orange circles* we used the number of SSPs that maximizes the evidence. The error bars correspond to the 16th and 84th percentile of the posterior distribution. The blue line corresponds to the IMF slope–velocity dispersion relation of Equation 3.13.

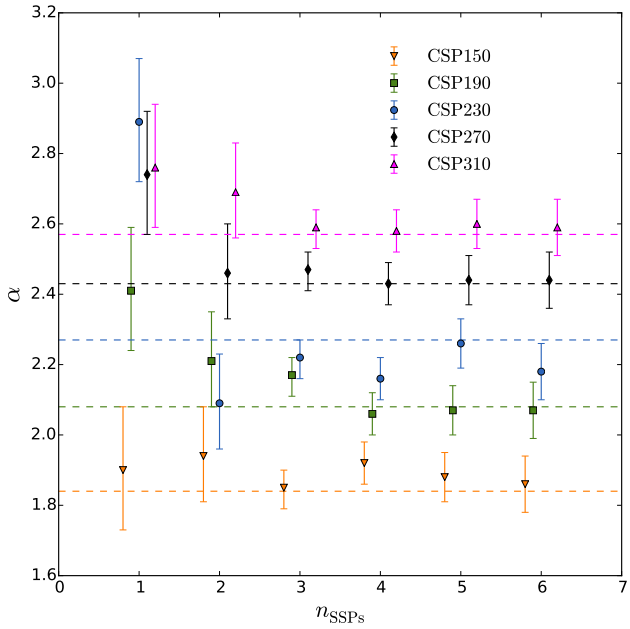

Figure 3.7 – Reconstructed IMF slopes as a function of the number of SSPs used in the fit for all mock spectra with SNR=150. The dashed lines correspond to the input values of the IMF slope used to create the mock spectrum. Different velocity dispersions are slightly offset to allow the reader to distinguish them.

3.5.4.1 IMF slope as a function of n_{SSPs}

We have seen that the evidence prefers fits with multiple SSPs over fits with only one SSP for all simulated spectra with realistic SFHs. Here we investigate in more detail if the determination of the IMF slope is affected by the number of SSPs that we use in our fit.

The black data points in Figure 3.6 show the IMF-slope–velocity dispersion relation that we find if we use only one SSP. Except for `CSP310-75`, the IMF slopes inferred by using only one SSP are biased to a higher value. The overall trend of an IMF slope that increases as a function of velocity dispersion is still recovered, however.

In Figure 3.7 we show the reconstructed IMF slope as a function of the number of SSPs used in the fit for all mock spectra with SNR=150. As we have seen in Figure 3.6, when we use only one SSP the reconstructed IMF slopes are biased and the slopes that we derive are in general too

high. For two SSPs there is still some bias, but for nearly all CSPs the input value of the IMF slope lies within the one sigma confidence interval of the reconstructed IMF slope. If we use three or more SSPs, the reconstructed IMF slopes converge to their true values, and the error bars on the reconstructed IMF slopes become approximately constant for more than three SSPs. So over a realistic range of SFHs, velocity dispersion (150-310 km s^{-1}) and SNRs (75-300), in general $N = 2 - 3$ SSPs are sufficient to recover the IMF without significant bias.

Given the wavelength region and the resolution of the data that we consider, we conclude that it should be sufficient to model CSPs with SFHs similar to the ones in Figure 3.1 with $\sim$3 SSPs. If we use only one SSP, this might significantly bias our reconstructed IMF slope. Although we still recover the trend, it is shifted upwards, indicating an overall overestimation of the IMF slope.

3.6 Summary and discussion

We have extended the hierarchical Bayesian framework of DTK16 to include multiple SSPs, a multiplicative polynomial and an extra covariance term. In addition, we have developed a parameterized version of the code which is much faster than the full Bayesian version. We use that version of the code to determine the velocity dispersion and the ages and metallicities of the SSPs that we fit to a spectrum. To reduce the number of templates in our model, we have implemented a binning procedure where we combine different isochrone stars.

We have applied the updated version of our model to a set of CSP mock spectra with different SNRs, stellar velocity dispersions and age-metallicity distributions. Based on the evidence, we determine the number of SSPs required to fit the mock spectra. The higher the SNR and the more extended the SFH of a CSP, the more SSPs are required to fit the spectrum correctly. As a rule of thumb, for SNR=75 there is strong evidence to include 2 SSPs in the fit, for SNR=150 there is strong evidence to include 2-3 SSPs in the fit and for SNR=300 there is strong evidence to include 5 SSPs in the fit. However, for removing the bias in the IMF slope $N = 2 - 3$ seems sufficient even for the high SNR = 300 cases.

Our CSP mock spectra have a single power law IMF with a slope that varies as a function of the present-day velocity dispersion. The assumed IMF slope–velocity dispersion relation is purely driven by observational

results. However, we do not expect this to be a physically realistic model since most of the stars in these galaxies formed long ago when the galaxy properties were quite different. Therefore, physically we expect that a variable IMF is driven by different local properties of the interstellar medium (ISM) at the time that the stars are born. These local properties of the ISM during star formation may typically be different for galaxies with different velocity dispersions/masses, which would then explain the observational results. In this work we use these observational results as a simple parameterization of the IMF but in principle within cosmological simulations one can relate the IMF to local conditions at the sites of star formation (see Blancato et al. 2016).

We show that we can reconstruct the underlying IMF slope–velocity dispersion relation for all the SNRs we consider. For this analysis, we fit the mock spectra with multiple SSPs and choose the number of SSPs that maximizes the Bayesian evidence. If we use only one SSP in our fit, our results show that the IMF slope that we determine is likely to be biased upward, although the trend remains. For three SSPs or more the reconstructed IMF slope appears to converge to a stable value but most of the bias is already removed by using two SSPs instead of one. In fact, Conroy & van Dokkum (2012b) use two SSP components in their fits, although the age of the second component is fixed to 3 Gyr. Martín-Navarro et al. (2015a) test the robustness of their inferred IMF gradient for NGC4552 by also modelling it with two SSPs instead of one and conclude that it makes no difference to the inferred gradient. Indeed, the inferred gradient remains if two SSPs are used instead of one, but the inferred IMF slopes appear to be lower, in particular for the outer radii. Our results suggest that IMF slopes derived on the basis of single SSP fits might be biased to a value that is too high if the SFH of the studied galaxy is in reality extended. Since high IMF slopes imply a higher stellar mass, this bias might help to explain discrepancies between stellar masses and masses derived through lensing and/or dynamics. However, despite this bias, any trends that are discovered appear to be real.

Although we have implemented an additional (constant) covariance term to absorb systematic uncertainties, this parameter does not seem to be significant here. The fact that our mock spectra have extended SFHs might potentially introduce the need to increase the covariance. However, this effect appears to be mostly absorbed by modelling the spectra with multiple SSPs. Note that if we model our spectra with one SSP only, the

additional covariance term ($b_{\rm cov}$) is in general significantly higher. Since the extended SFHs of our mock spectra are the only source of systematic uncertainty in this work, the fact that $b_{\rm cov}$ is in general relatively small is to be expected. If we deal with real data which is affected by other potential sources of uncertainty we expect this term to become more important, and potentially have a complex covariance matrix structure (e.g. due to correlations).

We model a CSP with a complex SFH by approximating it as the sum of a given number of SSPs. The aim of this approach is to investigate whether the inferred IMF is affected by the number of SSPs used in the fit. Although our method gives us some constraint on the SFH, it does not allow us to fully reconstruct the SFH of the CSP. One could imagine a data model similar to that present here that allows one to infer the SFH and the IMF simultaneously. In that case, the number of stellar templates in the matrix $\mathbf{S}$ would increase considerably, and we would need to include a parameterized prior on both the SFH and the IMF to regulate the inversion of Equation 3.1. However, that approach is computationally too expensive at present.

As shown in this work, the CSPs that we consider can effectively be modelled by a combination of a few SSPs, since there is a peak in the evidence as a function of the number of SSPs. Cosmological simulations predict that massive galaxies are made up of a significant fraction of stars that are formed in galaxies other then the main progenitor but are later on accreted through mergers (Naab et al. 2009; Porter et al. 2014b; Rodriguez-Gomez et al. 2016). This is one reason why these galaxies are expected to have composite stellar populations. In that context, it would be interesting to see if the SSPs that we recover as the main building blocks of a galaxy correspond to the most important constituent progenitors in the SAM. If so, this would allow us to put constraints on the merger history of the galaxy.

In this study our focus has been on the IMF. If instead one is interested in the SFH, one could assume an IMF and fill matrix $\mathbf{S}$ with a set of SSP templates instead of stellar templates. By using a parameterized prior on the SFH, one could then create a framework similar to our model that allows for hierarchical Bayesian inference of the SFH (see also Ocvirk et al. 2006, who use regularized inversion to determine non-parametric SFHs of CSPs).

Another possibility to combine inference of both the SFH and the IMF, would be to partition the three dimensional space of $T_{\rm eff}$, $\log g$ and [Fe/H] around the isochrones into small cubes. For each of these cubes, we could create a spectrum by using an interpolator and then fill matrix $\mathbf{S}$ with these spectra. The prior on the weights that we use to to regulate the inversion of Equation 3.1 now depends on both the parameterized SFH prior and the parameterized IMF prior. Although in principle this allows for a combined inference of the SFH and the IMF, this approach will still be very computationally expensive and may also be complicated by possible degeneracies between the parameterized SFH and the parameterized IMF. Note that for example Spinrad & Taylor (1971) and Faber (1972) already developed population synthesis models that tried to determine individual contributions of stars in the HR diagram. However, these methods combined stars in a rather ad hoc way and had few astrophysical constraints.

In this work we have shown that, given the wavelength range and resolution of the MILES library, we can reconstruct a variable IMF in CSPs. However, it is important to realize that we have only considered single power law parameterizations of the IMF prior. Based on dynamical constraints, Lyubenova et al. (2016) rule out a single power law IMF for the majority of their sample of 27 ETGs whereas it is found to be consistent with a double power law IMF for most of their sample. We might also consider more realistic parameterizations of the IMF, such as a double power law IMF with a break. If we would then create a set of CSP mock spectra where only the low-mass slope varies as a function of velocity dispersion, it could become much harder to reconstruct the IMF (variability). The contribution of low-mass stars to the spectrum is relatively small, and, as shown in DTK16, there is a degeneracy between the low-mass and the high-mass slopes. Since it is hard to distinguish dwarfs from giants in the wavelength range of the MILES library, we currently do not attempt more complicated parameterizations of the IMF prior. In future work, we plan to combine our model with the X-shooter Spectral Library (Chen et al. 2011). This empirical stellar library provides a good coverage of the HR diagram and has a wavelength range that extends from the UV to the NIR. The wavelength range of this library contains many more features that allow us to distinguish between dwarfs and giants. With an expanded wavelength range, we should also be able to probe more complicated parameterizations of the IMF prior.

Acknowledgements

We thank the anonymous referee for the useful comments which improved the quality of this paper. This work was supported in part by an NWO grant (project number 614.001.208) to SCT and by an NWO-VICI career grant (project number 639.043.308) to LVEK. RSS thanks the Downsbrough family for their generous support, and gratefully acknowledges support from the Simons Foundation through a Simons Investigator grant.

Appendix 3.A Binning isochrone stars

To reduce the computational costs of running our model, we bin the stellar templates of an SSP. This procedure is discussed in Section 3.3.3. When we combine different stellar templates into one new template, we need to assume an *a priori* IMF, since binning affects the structure of matrix **S** and cannot be done on-the-fly. In this paper we assume this a priori IMF in the binning procedure to have a Salpeter IMF slope inside the bin. This will slightly shift the average mass inside the bin. However, since the bins are very narrow, this assumption is of very low order. Here, we investigate the consequences of this assumption in more detail.

First, we determine the effect of our binning procedure on a set of SSP spectra. For every grid point in our age-metallicity grid, we create two SSP spectra: one created with the full set of stellar templates and one created with the binned set of stellar templates. Then, we calculate the residual between the two SSP spectra through

$$R_{\rm S} = \left\langle \frac{\mathrm{abs}\left(S_{\rm SSP} - S_{\rm SSP,bin}\right)}{S_{\rm SSP}} \right\rangle, \tag{3.17}$$

where $S_{\rm SSP}$ is the SSP spectrum created with the full set of stellar templates and $S_{\rm SSP,bin}$ is the SSP spectrum created with the binned set of stellar templates. Assuming a single power law IMF, we repeat this for different IMF slopes in the range $\alpha = [1.5, 3.1]$, where $\alpha = 2.35$ corresponds to a Salpeter IMF. The results of this analysis are shown in Figure 3.8. The residual between two SSP spectra is never more than 0.05%. As expected, the residual becomes smaller as the IMF slope is closer to Salpeter and for a Salpeter IMF the two SSP spectra are exactly the same.

Secondly, we test the effect of our binning procedure on the inference of the IMF. For this test, we have created six mock SSPs with different ages,

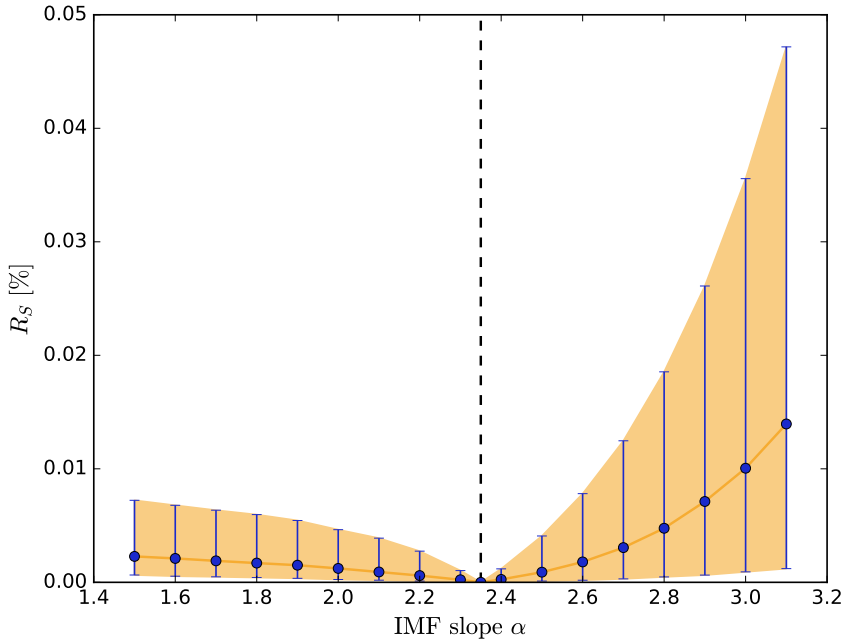

Figure 3.8 – Residuals between SSP spectra created with full set of stellar templates and SSP spectra created with binned version of stellar templates for different IMF slopes. The blue points correspond to the median residual of the age-metallicity grid for each of the IMF slopes that we consider. The error bars and the shaded orange region correspond to the 2σ values of the distribution of residuals. The black-dashed line represent the Salpeter IMF slope.

Table 3.3 – Parameters of six mock SSPs that we use to test if binning the stellar templates under the assumption of a Salpeter IMF affects our inference of the IMF slope.

name	age [Gyr]	[Fe/H]	α
`SSPt1`	3.0	0.21	1.7
`SSPt2`	3.0	0.21	3.0
`SSPt3`	8.0	-0.05	1.7
`SSPt4`	8.0	-0.05	3.0
`SSPt5`	12.0	-0.30	1.7
`SSPt6`	12.0	-0.30	3.0

Table 3.4 – Reconstructed parameters of the mock SSPs in Table 3.3.

		reconstructed		
name	age [Gyr]	[Fe/H]	σ	α
SSPt1	3.0	0.21	150.3	$1.66^{+0.04}_{-0.03}$
SSPt2	3.0	0.21	150.5	$3.03^{+0.02}_{-0.02}$
SSPt3	8.0	-0.05	150.3	$1.70^{+0.04}_{-0.04}$
SSPt4	8.0	-0.05	149.8	$2.99^{+0.02}_{-0.02}$
SSPt5	12.0	-0.30	150.2	$1.71^{+0.03}_{-0.04}$
SSPt6	12.0	-0.30	149.9	$3.00^{+0.01}_{-0.01}$

metallicities and IMF slopes. These mock SSPs have been created with the full set of stellar templates, have an average SNR of 300 and are smoothed to a velocity dispersion of $150\,\mathrm{km\,s^{-1}}$. The parameters of the SSPs are summarized in Table A1. Then, we reconstruct the IMF slopes of these SSPs with our model using the binned version of the stellar templates. The reconstructed parameters are given in Table A2 and show that the inferred IMF slope for these SSPs is not affected by our binning procedure. Note that we considered two IMF slopes which significantly deviate from a Salpeter IMF slope, because if there is any effect, the effect is expected to be largest for these extreme slopes. Based on these tests, we conclude that the effect of binning can be fully neglected at the current levels of signal-to-noise (being $\ll 2000$ to reach the level of 0.05% residuals).

Appendix 3.B Systematic uncertainties

To get a feeling for the effect of systematic uncertainties on the differences in log evidence that we find, we perform an additional test with three mock SSPs. The ages and metallicities of these mock SSPs are given in Table 3.5. All the mock SSPs in this Appendix have a Salpeter IMF and SNR = 150. First, we run the model with the complete set of stellar templates and calculate the evidence for the three mock SSPs. Then we remove the original stellar templates that were used to create the mock SSPs from the library of stellar templates (i.e. we remove all stellar templates with the age and metallicity of the mock SSP) and run the model again. In this way, we mock incompleteness of the basis set, which will give us some feeling

Table 3.5 – Evidence for three mock SSPs calculated with and without the original stellar templates that were used to create the mock SSPs.

name	age [Gyr]	[Fe/H]	evidence with original templates	evidence without original templates
SSPt7	3.0	0.21	10474.8	10450.5
SSPt8	8.0	-0.05	10506.4	10499.9
SSPt9	12.0	-0.30	10490.0	10486.5

on how to interpret differences in log evidence with respect to systematic uncertainties.

The results of the additional test are summarized in Table 3.5. For the youngest SSP, the difference in log evidence is relatively large. However, for the other two SSPs the difference in log evidence between running the model with and without the original stellar templates is well below ten. In all cases the model prefers to use the original set of stellar templates over the modified set of stellar templates. These results suggest that when we find a difference of log evidence that is more than $10^{3/2} \sim 32$, which according to Jeffreys scale is considered as very strong evidence, that this difference in log evidence is meaningful in the context of systematic uncertainties. We emphasize though that the result of this test is also strongly dependent on the adopted spacings of the age-metallicity grid. In Table 3.6 we show the inferred parameters of the SSPs that we obtain with the original stellar templates and with the modified stellar templates. The results that we infer with both sets of templates are consistent with each other within the errors, which implies that the results are robust against perturbations of the basis set of stellar templates.

Table 3.6 – Reconstructed parameters of the three mock SSPs considered in Table 3.5 with the original set of stellar templates and with the modified set of stellar templates.

name	reconstructed age [Gyr]	reconstructed [Fe/H]	reconstructed α
	original templates		
SSPt7	3.0	0.21	2.33 ± 0.04
SSPt8	8.0	-0.05	2.33 ± 0.04
SSPt9	12.0	-0.30	2.33 ± 0.03
	modified templates		
SSPt7	2.8	0.21	2.39 ± 0.28
SSPt8	8.3	-0.05	2.25 ± 0.09
SSPt9	11.8	-0.30	2.34 ± 0.05

References

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
Blancato, K., Genel, S., & Bryan, G. 2016, ArXiv e-prints
Brennan, R., Pandya, V., Somerville, R. S., et al. 2017, MNRAS, 465, 619
Brennan, R., Pandya, V., Somerville, R. S., et al. 2015, MNRAS, 451, 2933
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000
Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2012, Nature, 484, 485
Cenarro, A. J., Gorgas, J., Vazdekis, A., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 339, L12
Chabrier, G. 2003, PASP, 115, 763
Chen, Y., Girardi, L., Bressan, A., et al. 2014, MNRAS, 444, 2525
Chen, Y., Trager, S., Peletier, R., & Lançon, A. 2011, Journal of Physics Conference Series, 328, 012023
Clauwens, B., Schaye, J., & Franx, M. 2016, MNRAS, 462, 2832
Conroy, C. & van Dokkum, P. 2012a, ApJ, 747, 69
Conroy, C. & van Dokkum, P. G. 2012b, ApJ, 760, 71
Conroy, C., van Dokkum, P. G., & Villaume, A. 2017, ApJ, 837, 166
Czekala, I., Andrews, S. M., Mandel, K. S., Hogg, D. W., & Green, G. M. 2015, ApJ, 812, 128
Dries, M., Trager, S. C., & Koopmans, L. V. E. 2016, MNRAS, 463, 886
Faber, S. M. 1972, A&A, 20, 361
Faber, S. M. & French, H. B. 1980, ApJ, 235, 405
Feroz, F. & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, ArXiv e-prints
Ferreras, I., La Barbera, F., de la Rosa, I. G., et al. 2013, MNRAS, 429, L15
Fontanot, F., De Lucia, G., Monaco, P., Somerville, R. S., & Santini, P. 2009, MNRAS, 397, 1776
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41

Gorgas, J., Faber, S. M., Burstein, D., et al. 1993, ApJS, 86, 153
Graves, G. J. & Faber, S. M. 2010, ApJ, 717, 803
Graves, G. J., Faber, S. M., & Schiavon, R. P. 2009, ApJ, 698, 1590
Jeffreys, H. 1961, Oxford Univ. Press
Kaviraj, S., Khochfar, S., Schawinski, K., et al. 2008, MNRAS, 388, 67
Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102
Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, A&A, 501, 1269
Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2008, MNRAS, 385, 1998
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
Kroupa, P. 2001, MNRAS, 322, 231
Kroupa, P., Tout, C. A., & Gilmore, G. 1993, MNRAS, 262, 545
La Barbera, F., Ferreras, I., Vazdekis, A., et al. 2013, MNRAS, 433, 3017
Lang, P., Wuyts, S., Somerville, R. S., et al. 2014, ApJ, 788, 11
Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. 2004, A&A, 425, 881
Lyubenova, M., Martín-Navarro, I., van de Ven, G., et al. 2016, MNRAS, 463, 3220
MacKay, D. J. C. 1992, Neural Computation, 4, 415
Maraston, C. 2005, MNRAS, 362, 799
Marigo, P., Bressan, A., Nanni, A., Girardi, L., & Pumo, M. L. 2013, MNRAS, 434, 488
Martín-Navarro, I., Barbera, F. L., Vazdekis, A., Falcón-Barroso, J., & Ferreras, I. 2015a, MNRAS, 447, 1033
Martín-Navarro, I., Vazdekis, A., La Barbera, F., et al. 2015b, ApJL, 806, L31
McDermid, R. M., Alatalo, K., Blitz, L., et al. 2015, MNRAS, 448, 3484
Monachesi, A., Trager, S. C., Lauer, T. R., et al. 2012, ApJ, 745, 97
Naab, T., Johansson, P. H., & Ostriker, J. P. 2009, ApJL, 699, L178
Newman, A. B., Smith, R. J., Conroy, C., Villaume, A., & van Dokkum, P. 2016, ArXiv e-prints
Ocvirk, P., Pichon, C., Lançon, A., & Thiébaut, E. 2006, MNRAS, 365, 46
Popping, G., Somerville, R. S., & Galametz, M. 2016a, ArXiv e-prints
Popping, G., Somerville, R. S., & Trager, S. C. 2014, MNRAS, 442, 2398
Popping, G., van Kampen, E., Decarli, R., et al. 2016b, MNRAS, 461, 93
Porter, L. A., Somerville, R. S., Croton, D. J., et al. 2012, ArXiv e-prints
Porter, L. A., Somerville, R. S., Primack, J. R., et al. 2014a, MNRAS, 445, 3092

Porter, L. A., Somerville, R. S., Primack, J. R., & Johansson, P. H. 2014b, MNRAS, 444, 942
Rodriguez-Gomez, V., Pillepich, A., Sales, L. V., et al. 2016, MNRAS, 458, 2371
Rodríguez-Puebla, A., Behroozi, P., Primack, J., et al. 2016, MNRAS, 462, 893
Rosenfield, P., Marigo, P., Girardi, L., et al. 2016, ApJ, 822, 73
Salpeter, E. E. 1955, ApJ, 121, 161
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Schiavon, R. P. 2007, ApJS, 171, 146
Schiavon, R. P., Barbuy, B., Rossi, S. C. F., & Milone, A. 1997a, ApJ, 479, 902
Schiavon, R. P., Barbuy, B., & Singh, P. D. 1997b, ApJ, 484, 499
Serra, P. & Trager, S. C. 2007, MNRAS, 374, 769
Smith, R. J., Lucey, J. R., & Conroy, C. 2015, MNRAS, 449, 3441
Somerville, R. S. & Davé, R. 2015, ARA&A, 53, 51
Somerville, R. S., Gilmore, R. C., Primack, J. R., & Domínguez, A. 2012, MNRAS, 423, 1992
Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
Somerville, R. S., Popping, G., & Trager, S. C. 2015, MNRAS, 453, 4337
Somerville, R. S. & Primack, J. R. 1999, MNRAS, 310, 1087
Spiniello, C., Trager, S., Koopmans, L. V. E., & Conroy, C. 2014, MNRAS, 438, 1483
Spiniello, C., Trager, S. C., Koopmans, L. V. E., & Chen, Y. P. 2012, ApJL, 753, L32
Spinrad, H. & Taylor, B. J. 1971, ApJS, 22, 445
Tinsley, B. M. 1968, ApJ, 151, 547
Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 120, 165
Trager, S. C. & Somerville, R. S. 2009, MNRAS, 395, 608
Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, ApJ, 709, 1195
Trujillo-Gomez, S., Klypin, A., Primack, J., & Romanowsky, A. J. 2011, ApJ, 742, 16
van Dokkum, P., Conroy, C., Villaume, A., Brodie, J., & Romanowsky, A. 2016, ArXiv e-prints
Vazdekis, A., Cenarro, A. J., Gorgas, J., Cardiel, N., & Peletier, R. F. 2003,

MNRAS, 340, 1317
Vazdekis, A., Coelho, P., Cassisi, S., et al. 2015, MNRAS, 449, 1177
Wing, R. F. & Ford, Jr., W. K. 1969, PASP, 81, 527
Worthey, G., Faber, S. M., Gonzalez, J. J., & Burstein, D. 1994, ApJS, 94, 687
Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45

Chapter 4

Recovering variations in the Initial Mass Function of nearby Early-Type Galaxies with hierarchical Bayesian inference

— **M. Dries, S.C. Trager, L.V.E. Koopmans and C. Spiniello**

—

in collaboration with the XSL team

Abstract

There is mounting evidence that the initial mass function (IMF) in massive early-type galaxies (ETGs) is not the same as the Milky Way IMF. In this work we apply a hierarchical Bayesian framework for stellar population synthesis (SPS) to a set of stacked SDSS spectra of ETGs binned as function of their stellar velocity dispersion. We combine this Bayesian framework with the MIX library, a combination of two stellar libraries: MILES and the VIS arm of the X-shooter Spectral Library (XSL). When we apply the model to the data, we include multiple single stellar populations, a flexible IMF, and variable abundance patterns. We test different model ingredients (response functions, isochrones, regularization schemes and IMF parameterizations) that we compare on the basis of the Bayesian evidence. For a single power law IMF, we find an IMF slope that increases from $\alpha \sim 2.1$ for $\sigma = 170\,\mathrm{km\,s^{-1}}$ to $\alpha \sim 2.6$ for $\sigma = 320\,\mathrm{km\,s^{-1}}$ (with $\alpha = 2.35$ being a Salpeter IMF). These results are consistent with previous results that are derived with an independent code. When using a double power law IMF with a break at $0.5\,\mathrm{M_\odot}$, we find an almost constant high-mass IMF slope $\alpha_2 \sim 2.5$ and a low-mass IMF slope that increases from $\alpha_1 \sim 1.3$ to $\alpha_1 \sim 2.4$ as a function of velocity dispersion. In almost all cases we find that the double power law IMF is preferred over the single power law IMF. However, the uncertainties in the model, in particular the response functions of Na, make it hard to establish the absolute scale of the relation between IMF slope(s) and velocity dispersion or the exact shape of the IMF.

4.1 Introduction

The distribution of stellar masses resulting from a star formation event is described by the initial mass function (IMF) $\mathrm{d}N/\mathrm{d}M$. Salpeter (1955) introduced the concept of an IMF and determined that for the solar neighbourhood it could be described with a single power law $\mathrm{d}N/\mathrm{d}M \propto M^{-\alpha}$ between 0.4 $\mathrm{M}_\odot$ and 10 $\mathrm{M}_\odot$. More precise measurements of the IMF based on star counts later showed that the Milky Way IMF becomes flatter below $M \lesssim 0.5$ $\mathrm{M}_\odot$ and can be described by a broken power law (Kroupa et al. 1993; Kroupa 2001) or a lognormal distribution extended with a power law (Chabrier 2003). Note, however, that most studies measure the present-day mass function (PDMF) and not the IMF. The PDMF includes the effects of stellar evolution, the star formation history (SFH) of the considered object and sometimes dynamical evolution leading to stellar mass segregation.

Studies of the IMF in different environments of the Milky Way seem to indicate that the Galactic IMF is universal (Bastian et al. 2010). Beyond the Local Group, star counts are not possible any more and measurements of the IMF have to resort to more indirect methods. Most of these studies focus on early-type galaxies (ETGs) because these objects are relatively simple in terms of their SFH and less contaminated by ongoing star formation than late-type galaxies. Increasing evidence from spectroscopic studies suggests that in these ETGs the IMF becomes more bottom-heavy for galaxies with a higher mass/velocity dispersion (Conroy & van Dokkum 2012b; Spiniello et al. 2012; Ferreras et al. 2013; La Barbera et al. 2013) or metallicity (Martín-Navarro et al. 2015). These results are confirmed by dynamical and gravitational lensing studies that infer higher mass-to-light ratios for these ETGs than the Milky Way-value (Treu et al. 2010; Graves & Faber 2010; Cappellari et al. 2012; Conroy & van Dokkum 2012b; Lyubenova et al. 2016). However, in contrast to these results Smith et al. (2015) and Newman et al. (2016) have found a number of lensed ETGs that are consistent with the Milky Way IMF.

The IMF is a fundamental quantity in many astrophysical studies. Mass-to-light ratios of galaxies, chemical evolution and nucleosynthesis, the population of stellar remnants and the energy balance of the interstellar medium all depend on the assumed IMF, which is often considered to be universal and similar to the Milky Way. If the IMF is not universal, this may have important consequences for galaxy properties derived under the

assumption of a universal IMF, such as the SFH and the chemical evolution (Ferré-Mateu et al. 2013; Fontanot et al. 2017). It is therefore crucial to establish the shape of the IMF and its possible variations with other galaxy properties such as the stellar mass or velocity dispersion. If the IMF is not universal, this is expected to be related to a change in the local properties of the star forming interstellar medium (e.g., Larson 1998; Chabrier 2003). Since these properties may change as a function of time, the IMF of a single galaxy may be different for various episodes of star formation. If this is the case, the IMF (PDMF) that we measure is actually the superposition of the different IMFs that characterize the SFH of the galaxy.

Star formation is a complex physical process that involves gravity, turbulence, radiation and magnetic fields. Developing a theoretical model for this process is challenging, and therefore the origin of the IMF is far from being understood (Krumholz 2014; Offner et al. 2014). Several attempts have been made to relate the IMF with local properties of the star-forming interstellar medium such as the pressure or the Mach number (e.g Hennebelle & Chabrier 2008; Krumholz 2011; Hopkins 2012, 2013). The link between these local properties and the typical conditions of star formation might help to explain the observed non-universality of the IMF. Conversely, accurate measurements of the IMF can provide important constraints for star formation theories.

Inferring the IMF from the spectrum of an unresolved galaxy is not straightforward, however. To determine the IMF of an unresolved stellar population through spectroscopy, one has to model its spectrum through a stellar population synthesis model (SPS). SPS models have been developed to transform the observable properties of a galaxy into physical properties that are not directly observable such as its SFH, chemical composition (often described by a metallicity and abundance pattern) and IMF. The first models were developed by Tinsley (1968), and nowadays a variety of different SPS models are available (e.g. Bruzual & Charlot 2003; Le Borgne et al. 2004; Maraston 2005; Conroy & van Dokkum 2012a; Vazdekis et al. 2015; Dries et al. 2016). All stellar population models consist of at least three ingredients: stellar evolution models (i.e. isochrones), a spectral library of stars and an IMF.

Of particular interest when inferring the IMF through a SPS model is the contribution of low-mass dwarfs ($M < 0.5$ $M_\odot$) to the integrated spectrum. The relative contribution of these stars to an optical spectrum is at most $\sim$10% for a 13 Gyr old single stellar population (SSP) with a

rather extreme bottom-heavy IMF (i.e. a single power law with $\alpha = 3.0$), but for a Kroupa IMF it is limited to only a few percent (Dries et al. 2016, DTK16 hereafter). However, the total mass in these stars represents ∼27% and ∼47% for a Kroupa IMF and a Salpeter IMF, respectively, between 0.1 $M_\odot$ and 100 $M_\odot$, making these stars very important in dynamical and lensing studies (Treu et al. 2010; Graves & Faber 2010; Cappellari et al. 2012; Conroy & van Dokkum 2012b; Lyubenova et al. 2016).

The already small contribution of low-mass dwarfs to the integrated spectrum is complicated even further by their spectral similarity to K and M giants. However, a number of (gravity- and pressure-sensitive) spectral features are known to be sensitive to either dwarfs or giants (Wing & Ford 1969; Faber & French 1980; Gorgas et al. 1993; Worthey et al. 1994; Schiavon et al. 1997a,b; Cenarro et al. 2003; Schiavon 2007; Spiniello et al. 2014), and these features provide important constraints for inferring the IMF and separating the relative contributions from dwarfs and giants.

Most SPS models assume a parameterization of the IMF and then determine the most likely values of one or two IMF slopes that are the free parameters within that parameterization. These maximum-likelihood IMF slopes define the ratio between dwarf and giant stars. Strikingly different IMF parameterizations may, however, result in similar dwarf-to-giant ratios (Clauwens et al. 2016). Measuring IMF slopes within a given parameterization of the IMF therefore does not directly reveal the shape of the IMF (which is fixed a priori by the IMF parameterization). DTK16 and Conroy et al. (2017) have developed more direct approaches to measure the contribution of individual mass bins to the integrated spectrum. Nevertheless, the priors in these models still rely on a parameterization of the IMF, resulting in restrictions on the allowed range of IMF shapes.

Another ingredient of SPS models that can affect the inference of the IMF is the assumption that the spectra of ETGs can be modelled as an SSP, with all stars formed at a single epoch. The typical SFH of an ETG, however, is expected to be characterized by an initial peak representing an old stellar population followed by an extended tail with a small amount of more recent star formation, as observed by e.g. Kaviraj et al. (2008), Monachesi et al. (2012) and McDermid et al. (2015). Dries et al. (2018) have shown that when their model is applied to a set of composite stellar population (CSP) mock spectra and only one SSP is used in the fit, the inferred IMF slopes are biased to a value that is too high (i.e. contains too many low-mass stars), typically inferring a slope that is ∼0.2 larger than

its real slope. When the spectra are modelled by multiple SSPs, typically two to three, this bias disappears.

One of the advantages of the hierarchical Bayesian framework developed by DTK16 and extended by Dries et al. (2018) (DTKPS18 hereafter) is that it allows one to compare different model ingredients on the basis of the Bayesian evidence. In this paper we extend the work of DTK16 and DTKPS18 by including variable abundance patterns and apply the hierarchical Bayesian framework to a set of stacked *Sloan Digital Sky Survey* (SDSS) spectra of ETGs, as function of their stellar velocity dispersion. We use a variety of different model ingredients to infer the IMF from these spectra and compare different ingredients on the basis of the Bayesian evidence. Among these ingredients there are up to four SSPs to model the spectra, two different sets of isochrones, response functions of various elements, two different parameterizations of the IMF and two different regularization schemes. Moreover, we introduce a new stellar library in our model: the MIX library, which is a combination of the MILES library (Sánchez-Blázquez et al. 2006) and the VIS arm of the X-shooter Spectral Library (Chen et al. 2011).

In Section 4.2.1 we describe the hierarchical Bayesian framework developed by DTK16 and extended by DTKPS18. The application of the model to real data requires us to introduce two new features into the model: response functions for non-solar abundance ratios and local covariance structures, which we describe in Section 4.2.2. In Section 4.3 we introduce the MIX library and describe how we prepare the library for using it in the Bayesian SPS framework. The SDSS data to which we apply the model is described in Section 4.4. In Section 4.5 we report our results and conclusions.

4.2 Model description

We use the hierarchical Bayesian framework developed by DTK16 and DTKPS18. In this work we introduce two additional features, response functions for non-solar abundance ratios and local covariance structures (i.e. a varying pattern of random and systematic errors on the spectrum), that allow us to improve the modelling of real galaxy spectra. We will first provide a short summary of the original model and then describe the new features that we introduce.

4.2.1 Hierarchical Bayesian framework

Apart from the effects of extinction, the spectrum of a galaxy is essentially the sum of the spectra of the stars that it contains. Following DTK16 and DTKPS18, this is expressed through a linear system of equations,

$$\boldsymbol{g} = \mathbf{S}\,\boldsymbol{w}, \tag{4.1}$$

in which $\boldsymbol{g}$ is the spectrum of the galaxy, $\mathbf{S}$ is a matrix with a representative set of stellar spectra, and $\boldsymbol{w}$ is a vector with the number of stars for each of the templates in $\mathbf{S}$. For an SSP, the spectra in $\mathbf{S}$ are defined by an isochrone as a function of age t and metallicity [Fe/H], and an abundance pattern $\vec{\alpha}$, so that $\mathbf{S} = \mathbf{S}(t, [\mathrm{Fe/H}], \vec{\alpha})$. The IMF,

$$\xi(M) \equiv \frac{\mathrm{d}N}{\mathrm{d}M}, \tag{4.2}$$

and the vector $\boldsymbol{w}$ (weights hereafter) are related through

$$\xi(m_j) = \frac{w_j}{m_{\mathrm{high}} - m_{\mathrm{low}}}, \tag{4.3}$$

and vice versa through

$$w_j = \int_{m_{\mathrm{low}}}^{m_{\mathrm{high}}} \xi(M)\mathrm{d}M, \tag{4.4}$$

where w_j is the number of stars for template j and m_{low} and m_{high} are the boundaries of the mass bin corresponding to this template.

The basic idea of the model developed by DTK16 is to infer the IMF of a stellar population through a regularized linear inversion of Equation 4.1. Regularization is necessary to ensure positive weights and realistic stellar mass distributions in case the inversion is ill determined. In DTK16 the inversion of Equation 4.1 is regularized by imposing a prior on the IMF. The prior on the IMF is derived from an assumed IMF parameterization, e.g. a (single) power law IMF, that is characterized by a set of non linear parameters $p_{i,0}$ (e.g., slope and normalization for a single power law IMF). The IMF parameterization can be translated into a prior on the weights $\boldsymbol{w}_0(p_{i,0})$ through Equation 4.4. Combining the likelihood with the prior on

the weights, Equation 4.1 can now be solved by minimizing

$$M(\boldsymbol{w}|\boldsymbol{g}, \mathbf{S}, \boldsymbol{w}_0, \mathbf{C}_{\rm pr}^{-1}) = \frac{1}{2}(\mathbf{S}\,\boldsymbol{w} - \boldsymbol{g})^{\rm T}\mathbf{C}_{\rm D}^{-1}(\mathbf{S}\,\boldsymbol{w} - \boldsymbol{g}) + \lambda \cdot \frac{1}{2}(\boldsymbol{w} - \boldsymbol{w}_0)^{\rm T}\mathbf{C}_{\rm pr}^{-1}(\boldsymbol{w} - \boldsymbol{w}_0), \tag{4.5}$$

where $\mathbf{C}_{\rm D}^{-1}$ is the inverse noise covariance matrix and $\mathbf{C}_{\rm pr}^{-1}$ is the inverse covariance matrix describing the shape of the Gaussian-distributed deviations of the weights from the prior. The first and second terms in Equation 4.5 correspond to, respectively, the likelihood and the prior and are balanced by the regularization parameter λ. The regularization parameter is chosen such that it optimizes the posterior probability but with an additional prior that prevents negative weights. If the IMF indeed follows the exact prior shape (for example a single power law), then the regularization penalty in Equation 4.5 goes to zero and the equation reduces to the usual fully parametric maximum-likelihood case.

The non-linear parameters related to the IMF prior and the age, metallicity and abundance pattern of the SSP, which define the stellar templates in $\mathbf{S}$, can be sampled via Markov Chain Monte Carlo (MCMC) sampling techniques. For every sample of these parameters, we calculate the most probable weights by minimizing Equation 4.5, and different samples of these parameters can be compared on the basis of the Bayesian evidence (MacKay 1992). At the second level of inference (model comparison), we can marginalize over the weights, ages, metallicities and IMF prior parameters and compare different models on the basis of the evidence (e.g. different IMF parameterizations or different isochrones).

In DTKPS18 the model of DTK16 was updated and extended with a number of features that we also include in this work. These additional features are allowing for multiple SSPs when fitting a spectrum, including the velocity dispersion as a free parameter in the model, the inclusion of a multiplicative polynomial in the fit and an adaptive covariance matrix. Since these additional features increase the computational time-scales significantly, DTKPS18 have also developed a parameterized version of the code that does not include a linear inversion of the weights but instead compares the prior spectrum $\boldsymbol{g}_0 = \mathbf{S}\boldsymbol{w}_0$ directly to the data $\boldsymbol{g}$ by varying the parameters that govern the IMF shape. This version of the code is much faster than the original version. When we infer the parameters related to a certain spectrum, we first use the parameterized version of the code to

determine the velocity dispersion and the most likely ages and metallicities of the SSPs included in the fit. Then, we fix the velocity dispersion and the ages and metallicities of the SSPs and run the full version of the model that also optimizes for the shape of the IMF and allows it deviate from the shapes of the prior model. Both the parameterized and the full version of the code are implemented as a pipeline of `cosmoSIS` (Zuntz et al. 2015). Evidences are calculated with `Multinest` (Feroz & Hobson 2008; Feroz et al. 2009, 2013). `Multinest` also provides a posterior sample which, if necessary, is refined with `emcee` (Foreman-Mackey et al. 2013).

For more details with respect to the hierarchical Bayesian framework and the additional features of the code we refer the reader to DTK16 and DTKPS18.

4.2.2 New features: abundance and error patterns

In Section 4.5 we apply our model to a set of stacked SDSS spectra. The galaxies corresponding to these spectra are mostly ETGs. ETGs are expected to have different abundance patterns than the Milky Way and are known to be alpha-enhanced (e.g. Trager et al. 2000a; Johansson et al. 2012; Conroy et al. 2014). These differences in abundance patterns can potentially affect the inference of the IMF (Conroy & van Dokkum 2012a). To account for these differences, we introduce response functions for a number of elements in the model. Since we are now dealing with real data, there are certain regions of the spectrum that are less reliable (e.g. regions affected by strong telluric absorption lines). Therefore, we include local covariance structures that allow the model to down-weight less reliable parts of the spectrum.

4.2.2.1 Response functions

Response functions describe the relative change of a spectrum as a consequence of a change in the chemical abundance pattern. In principal, response functions of a given element can affect the entire spectrum and not only the lines of that element. We use the response functions of Conroy & van Dokkum (2012a) and Conroy et al. (submitted). These response functions[1] are based on the ATLAS model atmosphere and spectrum synthesis package (Kurucz 1970, 1993) and are derived for various elements. The re-

[1]Downloaded from `https://github.com/cconroy20/alf/tree/master/infiles`

sponse functions are available for SSPs with ages $t = \{1.0, 3.0, 5.0, 9.0, 13.0\}$ Gyr, metallicities [Fe/H] $= \{-1.5, -1.0, -0.5, 0.0, 0.2\}$ and Salpeter and Kroupa IMFs.

To prepare the response functions for use in the models we first rebin them to the wavelength scale of the data. Then we linearly extrapolate the response functions to $t = 13.5$ Gyr and [Fe/H] $= 0.4$. This ensures that the ages and metallicities of the isochrones that we consider (see Section 4.3.5) are located within the age-metallicity grid of the response functions, and we can use bilinear interpolation to interpolate between the response functions. For every element considered, there is a response function that corresponds to an increase of 0.3 dex in the abundance of that element and a response function that corresponds to a decrease of 0.3 dex. Every age-metallicity combination of the isochrones corresponds to an SSP, and for each of these combinations we use bilinear interpolation to create two response functions (increase/decrease) for every element considered. When we include the response function of a given element X in the model, we sample the abundance variation [X/Fe] in dex as an additional free parameter and use linear interpolation between a response function that is a straight line at 1.0 (i.e. no change in the model spectrum) and the response function at +/-0.3 dex to account for an increase/decrease of [X/Fe]. Subsequently, the stellar templates in $\mathbf{S}$ are multiplied with the response functions of the corresponding age-metallicity interpolated to a value of [X/Fe] dex. If we use multiple SSPs, we use a different response function for every SSP. Since the computational time scales for running the model are already significant, we want to minimize the number of extra parameters that we have to sample. Therefore, we currently only consider the response functions of the alpha elements Mg, Ca, Ti and Si and the response function of Na. The models from which the response functions are derived are smoothed to $100\,\mathrm{km\,s^{-1}}$. Therefore, the spectra of the stellar templates in $\mathbf{S}$ are also smoothed to $\sigma = 100\,\mathrm{km\,s^{-1}}$. Note that we use the parameter [X/Fe] to quantify the relative strength of the response functions, we do not aim to use this parameter to establish the absolute scale of the abundance variations.

The response functions that we use in this work were calculated for SSPs with a Salpeter IMF. This is a simplification of a more complicated problem. In principle, to correctly account for abundance variations one needs to apply response functions to the individual stellar templates in matrix $\mathbf{S}$. However, these individual stellar response functions are not available to us,

and therefore this approach is not feasible in the current work but might be implemented in a future update of the code. The average difference between response functions derived under the assumption of a Salpeter IMF and response functions derived under the assumption of a Kroupa IMF are, however, in general relatively small for Mg, Ca and Si although significantly higher for Ti and Na. These differences are also higher for higher metallicities.

4.2.2.2 Additional local covariance

When we model the spectrum of a galaxy, certain regions of the spectrum are characterized by higher uncertainties due to e.g. telluric absorption. Moreover, the stellar templates in the model currently do not allow us to correctly account for emission lines in the spectrum. We model these additional uncertainties by introducing two local covariance structures as a function of wavelength: a boxcar covariance structure and a Gaussian covariance structure. The boxcar structure can be used to add local covariance to strong telluric regions, while the Gaussian structure can be used to add local covariance to emission line regions. Both structures are characterized by three free parameters. For the boxcar these parameters are the amplitude of the additional noise, the central wavelength and the width over which the variance is allowed to vary, while for the Gaussian the width is replaced by the standard deviation. In principle we can include these as free parameters in the model, but this will increase the computational time scales significantly. Therefore, we currently choose not to vary the local covariance structures and add them only if necessary.

4.3 The MIX stellar population models

The MIX stellar population models are based on the MIX library, a combination of the MILES library (Sánchez-Blázquez et al. 2006) and the X-shooter Spectral Library (Chen et al. 2011, Lyubenova et al., in prep.). Here we describe the construction of the MIX stellar population models. In Section 4.3.1 we describe how we prepare the X-shooter Spectral Library (XSL) for building population synthesis models. Then we describe how we combine XSL and MILES into the single MIX library in Section 4.3.2. Since there are relatively few M dwarfs in XSL, we extend XSL with an additional set of M dwarfs from the ESO X-shooter archive which we discuss

in Section 4.3.3. In Section 4.3.4 we discuss how we correct the MIX library for uncertainties in the extinction and the stellar parameters by an internal optimization procedure. Finally, in Section 4.3.5 we describe how we combine the MIX libary with two different sets of isochrones to form the MIX stellar population models.

4.3.1 The X-shooter Spectral Library

XSL is a library of stellar spectra observed with the X-shooter spectrograph (Vernet et al. 2011) on ESO's VLT. The X-shooter spectrograph consists of three arms (UVB, VIS and NIR) and allows one to take a spectrum from the ultraviolet to the near-infrared (∼300-2500 nm) in a single exposure. Spectra in the XSL have a moderate resolution of ∼10,000 and are selected to provide a good coverage of the Hertzsprung-Russell (HR) diagram. The data reduction and calibration of XSL are described in Lyubenova et al. (in prep.). Currently, the library contains 815 spectra of 679 unique stars. These spectra are flux-calibrated and corrected for telluric absorption.

Before we can use the spectra for population synthesis models, we have to shift the spectra to their rest frame and to correct for Galactic extinction. The effect of Galactic extinction is stronger in the UVB arm than it is in the VIS arm. Moreover, there is a significant number of cool stars for which the signal-to-noise ratio (SNR) in the UVB is relatively low. When we apply an extinction correction to these spectra, this will amplify the noise in the spectra and potentially affect the inference of the parameters that we derive on the basis of these models. Apart from the extinction correction, the response curve in the overlap region between the UVB and the VIS arm (about 5500-6000 Å) shows features that are related to the dichroic that splits the incoming light between the two arms. Therefore, it is not straightforward to combine the spectra of the UVB and the VIS arm.

Considering the problems in the overlap region between the UVB and the VIS arm and the uncertain extinction correction in the UVB arm, we choose not yet to use the UVB arm in this work. Instead, we extend the VIS arm to the blue by using the MILES stellar library. Note that we do not need the NIR arm for the analysis in this work because the SDSS spectra only extend to ~ 9200 ÅIn the rest of this section we discuss the stellar parameters, the rest frame correction and the extinction correction of the spectra in the VIS arm.

4.3.1.1 Stellar parameters and radial velocities

The stellar parameters (effective temperature $T_{\rm eff}$, surface gravity $\log g$ and metallicity [Fe/H]) of the stars in XSL are determined by Arentsen et al. (in prep.). These stellar parameters are determined using ULySS (Koleva et al. 2009) with the MILES interpolator of Prugniel et al. (2011) and the improvements on this interpolator by Sharma et al. (2016). Arentsen et al. (in prep.) determine two sets of stellar parameters: the first set is determined from the UVB arm in the wavelength range 4000-5500 Å and the second set is determined from the VIS arm in the wavelength range 6000-7400 Å. The medians of the absolute differences between the two sets of stellar parameters are $\Delta \log T_{\rm eff} = 0.007$, $\Delta \log g = 0.17$ and Δ[Fe/H] = 0.10. In general, the stellar parameters derived from the UVB arm are considered to be more reliable than those derived from the VIS arm. However, for cooler stars there can be little flux in the UVB arm, resulting in low SNRs. Therefore, we use the parameters derived from the VIS arm below $T_{\rm eff} = 4000$ K, and the parameters derived from the UVB arm for the other spectra.

As a by-product of the determination of the stellar parameters, Arentsen et al. (in prep.) also derive radial velocities of the stars in the UVB and the VIS arms. We use the radial velocities derived from the VIS arm to shift the spectra in the VIS arm to rest frame. For 49 spectra there are no stellar parameters and/or radial velocities available. We do not use these spectra and in total 766 spectra remain for the construction of the MIX library. Figure 4.1 shows the HR diagram of all the XSL stars for which parameters have been determined. The stars are coloured by metallicity.

4.3.1.2 Extinction correction

To correct the stars in XSL (in this work only the VIS spectra) for Galactic extinction we assume a Cardelli extinction curve (Cardelli et al. 1989) with $R_V = 3.1$. For a given value of the extinction in the V-band A_V, the Cardelli extinction law allows us to calculate the extinction curve $A(\lambda)$ that we can use to de-redden our observations.

Extinction can be highly variable and changes as a function of position on the sky. Although extinction maps exist that provide the extinction for a given line of sight (Schlegel et al. 1998; Schlafly & Finkbeiner 2011), the extinction also depends on the distance. In that respect it would be better to use 3D extinction maps (Sale et al. 2014), but this requires accurate

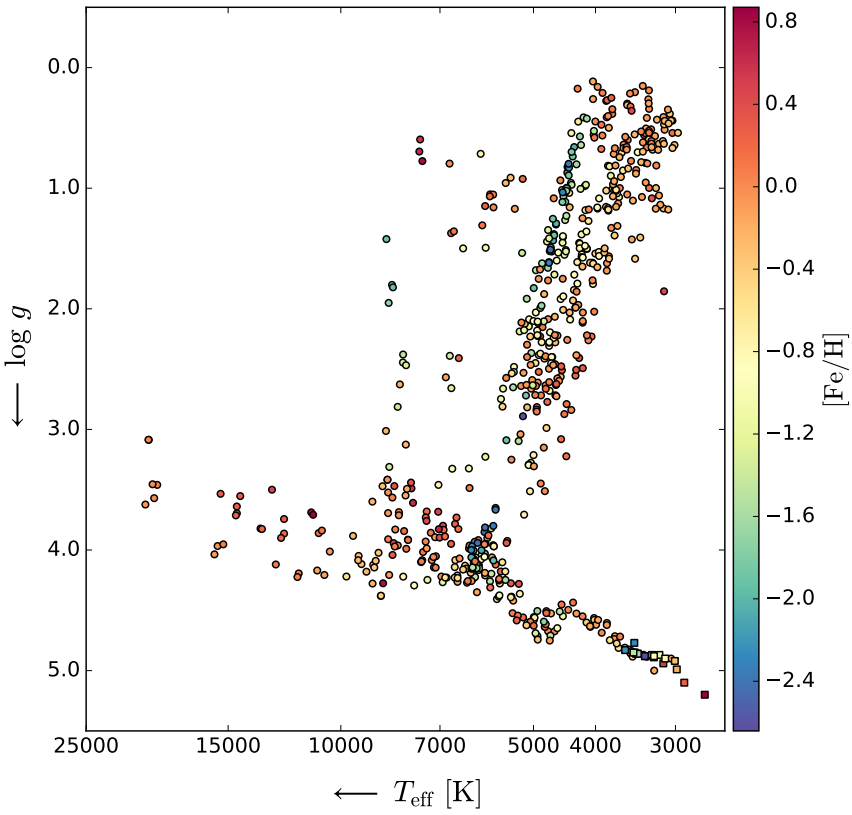

Figure 4.1 – HR diagram of the XSL stars with the (original) parameters of Arentsen et al. (in prep.). The stars are coloured by metallicity. The circles correspond to the stars in XSL and the squares to the additional M-dwarfs from the ESO X-shooter archive.

knowledge of the distance to the observed object. Since we do not have this information for all our stars and 3D extinction maps are not available for all directions, we do not use extinction maps to determine the values of A_V for the XSL stars. Instead, we calibrate the spectra against the MILES library to find the A_V-values. As we will later on combine XSL and MILES into the MIX library, this is a natural choice that ensures consistency between the two libraries although there might overall still be some uncertainties in the extinction correction.

The stars in XSL are partly chosen to overlap with the MILES library. For the 197 spectra that XSL has in common with the MILES library, we

use the extinction values derived for the MILES stars in Sánchez-Blázquez et al. (2006) to de-redden the spectra. We compare the remaining spectra against an interpolated MILES spectrum with the stellar parameters of the XSL spectrum. The interpolated spectrum is created with the interpolator described in DTK16 without the additional polynomial correction. We then optimize A_V by minimizing the residual between the XSL spectrum and the interpolated MILES spectrum in the wavelength range 6000-7000 Å. We use the resulting optimal value of A_V to de-redden the XSL spectrum.

Note that the A_V-values that we derive in this section are only used as starting values that allow us to build a first version of the MIX interpolator. In Section 4.3.4 we describe how we optimize this interpolator internally and derive new values of A_V to build a consistent interpolator.

4.3.2 The MIX library

The spectra in the MIX library are based on the VIS spectra of XSL which are extended towards the blue by using interpolated MILES spectra. We smooth the VIS arm spectra of XSL with a Gaussian kernel to a velocity dispersion of $\sigma = 100\,\mathrm{km\,s^{-1}}$ to match the resolution of the response functions. The original resolution of the VIS arm is assumed to have a constant value of 10,640 (FWHM)[2]. For each of the XSL spectra, we interpolate a MILES spectrum with the stellar parameters of the XSL star using the MILES interpolator of DTK16 (without the polynomial correction). The XSL parameters are based on the ULySS MILES interpolator and should therefore be consistent with the MILES parameters. The interpolated MILES spectra are also smoothed to a velocity resolution of $\sigma = 100\,\mathrm{km\,s^{-1}}$ with a (variable) Gaussian kernel. Then we divide the wavelength range of the final MIX spectrum into three regions. Below 6000 Å we use the interpolated MILES spectrum, between 6000 and 6500 Å we use linear interpolation between the interpolated MILES spectrum and the VIS spectrum and above 6500 Å we use the VIS spectrum. In the overlap region between 6000 and 6500 Å the weight of the interpolated MILES spectrum decreases linearly from one to zero whereas the weight of the VIS spectrum increases linearly from zero to one.

The wavelength range of the MIX library and the three different regions that we distinguish are summarized in Table 4.1. The MIX library contains 815 spectra of which 766 with stellar parameters and radial velocities.

[2]https://www.eso.org/sci/facilities/paranal/instruments/xshooter/inst.html

Table 4.1 – Overview of which spectra are used in the different wavelength regions of the MIX library.

$\lambda_{\rm start}$	$\lambda_{\rm end}$	spectrum
3540.5	6000	interpolated MILES spectrum
6000	6500	linear interpolation between interpolated MILES spectrum and VIS spectrum
6500	10000	VIS spectrum

However, we a priori remove 69 spectra from the library that are not suitable for interpolation (carbon stars, spectra with emission lines and spectra with low SNRs). This leaves a total of 697 spectra that we use as stellar templates. To these we will still add some additional M dwarf spectra.

4.3.3 Additional M dwarfs

The stars that are present in XSL do not provide a good coverage of the main sequence M dwarfs. Therefore, we extend the VIS arm of XSL with a number of additional M dwarfs from the ESO X-shooter archive. These additional M dwarfs are selected from ESO programs 385.D-0200 (chromospheric structure of low-mass stars, PI A. Reiners), 088.D-0556 and 092.D-0300 (observations of M dwarf secondaries, PI V. Neves). We only select spectra for which we can find literature values for their effective temperature and their surface gravity. This leaves a total of 23 additional M dwarfs. We use these literature parameters as an initial estimate for the optimization procedure described in Section 4.3.4. For their extinction we assume an initial guess of $A_V = 0.0$ motivated by the fact that these stars are faint and therefore are expected to be relatively close to the Solar System.

After downloading the raw data from the ESO archive, we reduce the spectra with the same pipeline as the other XSL stars. However, for the archive M dwarf spectra there are no wide-slit spectra available, and therefore we have to use a different approach for the flux calibration. In particular, we use the SDSS M dwarf templates provided by Bochanski et al. (2007). First, we shift the spectra of the XSL M dwarfs to the rest frame and smooth the spectra to the resolution of SDSS. Then we select the best-fitting SDSS template from Bochanski et al. (2007) where we allow

for a third-order multiplicative polynomial in the fit. After selecting the best-fitting SDSS template, we correct the XSL M dwarf with a third-order polynomial to absorb the continuum mismatch with the template. We expect that this correction (at least partially) removes issues with the continuum of the M dwarf related to the flux calibration and extinction.

As a final step, the reduced and flux calibrated VIS spectra of the M dwarfs from the archive are combined with an interpolated MILES spectrum as described in Section 4.3.2. With these additional M dwarfs the MIX library now contains 720 spectra. The improved coverage of the low-mass main-sequence is clearly shown in Figure 4.1, which shows the HR diagram of XSL together with the additional M dwarfs from the archive.

Although the MILES library also suffers from a poor coverage of the M dwarf region, we only use the interpolated MILES spectrum below 6500 Å. Since M dwarfs have much more flux in the red/NIR part of the spectrum, the flux level in the MILES wavelength region is expected to be relatively small. Therefore we are confident that the small number of M dwarfs in the MILES library is not a problem for the MIX library.

4.3.4 Optimizing the MIX library

By combining the extinction-corrected rest frame spectra of the VIS arm of XSL with an interpolated MILES spectrum below 6500 Å, we have created a first version of the MIX library. We use these spectra to build an initial version of the MIX interpolator, which is described in Section 4.3.4.1.

Uncertainties in the extinction-correction values and the stellar parameters that we use affect the quality of the interpolator. Therefore, we apply an internal optimization procedure with the extinction and small deviations in the stellar parameters as free parameters. The aim of this optimization procedure is to build a consistent interpolator. We describe the optimization procedure and the quality of the interpolator in Section 4.3.4.2.

4.3.4.1 MIX interpolator

A spectral interpolator can be used to interpolate between stars in a stellar library and to create a spectrum for an arbitrary set of stellar parameters. For the MIX interpolator we use almost the same interpolation algorithm as in DTK16. This is a local interpolation algorithm based on Vazdekis et al. (2003). Basically, in the three-dimensional space of θ (with $\theta = 5040/T_{\rm eff}$),

$\log g$ and [Fe/H], the algorithm first selects stars in the eight cubes that surround the requested interpolation point. Then, a weighting scheme is applied to combine the spectra of the selected stars into an interpolated spectrum. For more details on the interpolation mechanism, we refer the reader to DTK16 and Vazdekis et al. (2003).

There are two differences though between the interpolator in DTK16 and the MIX interpolator. First, the interpolator of DTK16 uses the MILES library, and here we apply the interpolation algorithm to the MIX library. Therefore, we change the maximum density of stars in the grid $\rho_{\rm M}$ (see DTK16) to the value of the MIX library. Second, we do not weigh the spectra by their SNR because we do not have a single SNR-value for the combined XSL and (interpolated) MILES spectrum.

4.3.4.2 Optimization procedure

As a next step we check the quality of the interpolator. To do this we compare the original spectrum with the interpolated spectrum for each star in the library. The star for which we make this comparison is not part of the dataset of the interpolator. We define the residual R_S as

$$R_S = \frac{\mathrm{abs}(\mathrm{S_{or}} - \mathrm{S_{int}})}{\mathrm{S_{or}}}, \tag{4.6}$$

with $\mathrm{S_{or}}$ the original spectrum and $\mathrm{S_{int}}$ the interpolated spectrum. When we determine the residual, we exclude strong telluric regions. For the first version of the MIX interpolator, the average residual is 3.4%. This is a relatively high average residual compared to the signal that we aim to detect[3]. In part this is related to uncertainties in the extinction values and the stellar parameters that we use. Therefore, we apply an internal optimization procedure where we allow for variations in these parameters.

The internal optimization procedure consists of two iterations. Since we consider the extinction to be more uncertain than the stellar parameters, in the first iteration we only optimize the extinction. We start by selecting the star with the largest residuals and we remove this star from the interpolator dataset. Then we optimize the extinction A_V by minimizing the residual

[3]According to the MIX models, for an old SSP (12.6 Gyr) the relative contribution of low-mass stars ($M \leq 0.5$ $\mathrm{M_\odot}$) to the integrated spectrum is between 2.5 and 4.0% for an IMF that varies between Kroupa and Salpeter. This number becomes lower if there are younger components that contribute to the integrated flux due to residual star formation in the considered object.

R_S between the original and the interpolated spectrum. Note that when we change A_V, this only affects the VIS spectrum of XSL while the interpolated MILES spectrum is left unchanged. We repeat this procedure for all stars in the library from the star with the highest residual to the star with the lowest residual. Each of the stars is only optimized once.

The second iteration is very similar to the first iteration, but now we also allow for small deviations in the stellar parameters. Since we optimize four parameters at once and there may be degeneracies between them, we use an MCMC approach to find the most likely combination of parameters. This is a rather sensitive procedure because changing the stellar parameters of one star affects the interpolated spectra of the surrounding stars as well. Therefore, we only allow for small deviations in the stellar parameters limited to $\Delta \log T_{\rm eff} = 0.02$, $\Delta \log g = 0.10$ and Δ[Fe/H] $= 0.2$. We only change the parameters of one star at a time. After optimizing $\sim$70 stars, the average residual converges to a value that is stable. Since we prefer to use the original parameters as much as possible, we stop the second iteration after 70 stars have been optimized. Figure 4.2 shows which stars in the HR diagram of XSL have been optimized. The majority of the optimized stars are found in the upper part of the HR diagram, a region characterized by uncertain and variable phases of stellar evolution. In this region, it may not be sufficient to interpolate stars on the basis of only three parameters (see also Mouhcine & Lançon 1999; Lançon & Wood 2000; Lançon & Mouhcine 2002). Therefore, the stellar parameters derived in this region may not be accurate. Potentially this introduces a bias on the interpolated spectra in this region but the optimization procedure at least ensures that the interpolator is consistent.

Apart from the 70 stars for which we optimize the stellar parameters and extinction in the second iteration, we have identified 26 spectra (including four of the additional M dwarfs) for which there was still a significant residual between the original spectrum and the interpolated spectrum after the optimization. We removed these spectra from the library to ensure that they do not affect the quality of the interpolator. Including the additional M dwarfs, the total number of spectra in the optimized MIX library is 694.

After applying the optimization procedure described in this section, the average residual between the original spectra and the interpolated spectra as defined in Equation 4.6 is 1.7%. In Figure 4.3 we show two examples of a comparison between the original spectrum of a star and its interpolated spectrum.

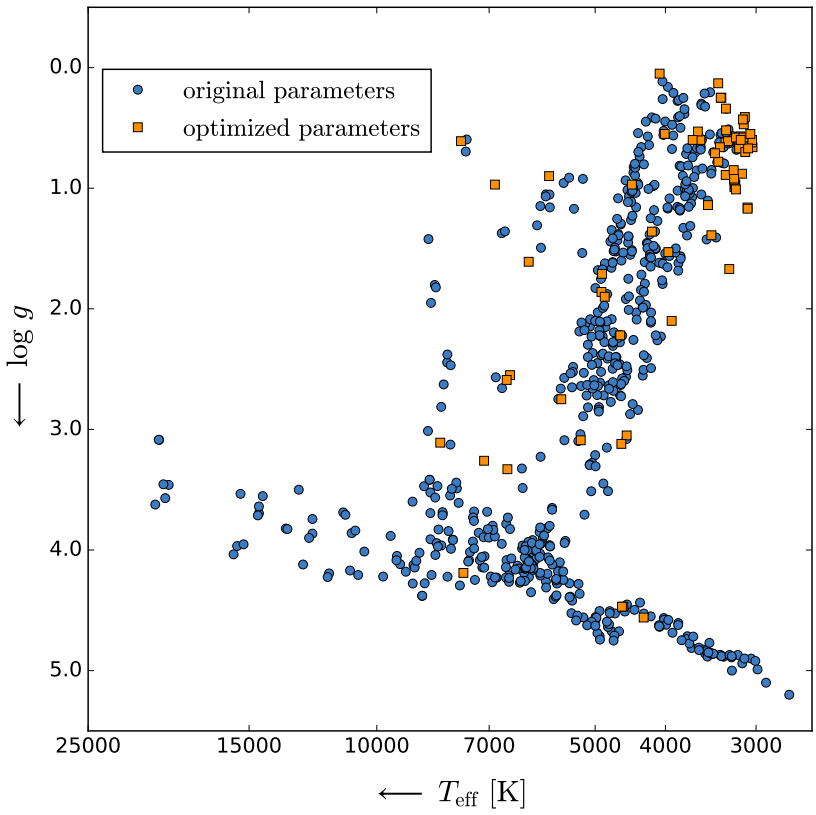

Figure 4.2 – Result of the optimization procedure of the MIX library. In addition to optimizing the extinction, we use the MIX interpolator to improve the stellar parameters of 70 stars. Only small variations in these parameters are allowed. The stars for which we optimized the stellar parameters are indicated by *orange squares*.

4.3.5 Stellar templates

For the analysis in this work we use two different sets of isochrones: the Parsec isochrones (Bressan et al. 2012) and a set of stitched isochrones very similar to the isochrones described in Conroy & van Dokkum (2012a). The stellar parameters ($\log T_{\rm eff}$, $\log g$ and the Johnson V-band magnitude) of these isochrones define the stellar templates of our model. We use the most recent version of the Parsec isochrones, which includes the thermally pulsating-asymptotic giant branch (TP-AGB) tracks from Marigo et al.

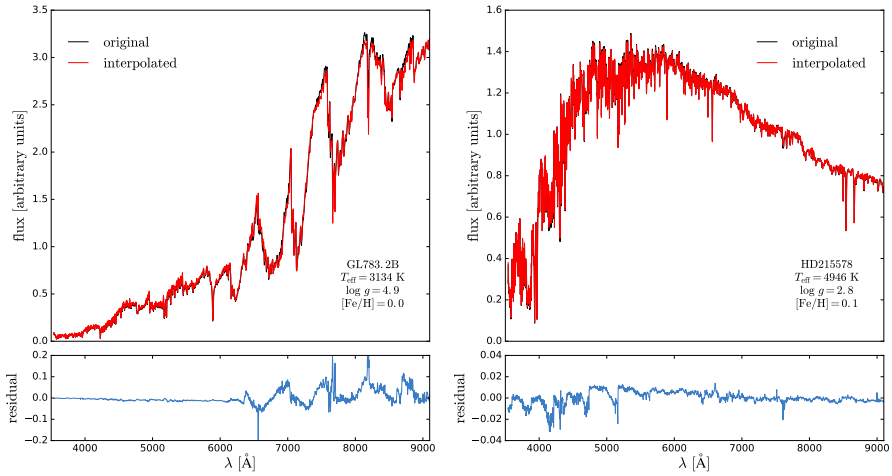

Figure 4.3 – Comparison between the original spectrum (*black*) and the interpolated spectrum (*red*) of two stars in the MIX library. The parameters of the star in the left panel (one of the additional M-dwards) are $T_{\rm eff} = 3134$ K, $\log g = 4.9$ and [Fe/H] = 0.0 while the parameters of the star in the right panel are $T_{\rm eff} = 4946$ K, $\log g = 2.8$ and [Fe/H] = 0.1. The residual between the original spectrum and the interpolated spectrum is shown in the bottom panels in *blue*.

(2013) and Rosenfield et al. (2016) and the improved treatment of low-mass stars described in Chen et al. (2014). The stitched isochrones that we use in this work are based on Conroy & van Dokkum (2012a) and have been developed with the idea in mind that there is no single set of isochrones that is optimized for all phases of stellar evolution. For $M < 0.17$ $M_\odot$ we use the Lyon isochrones derived from the NextGen models (Baraffe et al. 1998; Hauschildt et al. 1999). Between $M = 0.17$ $M_\odot$ and the tip of the red giant branch (RGB) we use the Dartmouth isochrones and for the horizontal branch (HB) and the asymptotic giant branch (AGB) we use the Padova isochrones (Marigo et al. 2008).

The stellar parameters of low-mass stars change very little as a function of age. Therefore, we use the average values of $\log T_{\rm eff}$, $\log g$ and V magnitude as a function of age for the Lyon isochrones. Assuming a low-mass cut-off (LMCO) of 0.1 $M_\odot$ (see e.g. Barnabè et al. 2013) this gives us stellar parameters for stars with $M = 0.1, 0.11, 0.13$ and 0.15 $M_\odot$ for all the isochrones (these mass data points are the same for all isochrones between 0.1 $M_\odot \leq M \leq 0.17$ $M_\odot$). The Lyon isochrones are defined for metallicities of [Fe/H] $= -2.0, -1.5, -1.3, -1.0, -0.5$ and 0.0.

For intermediate metallicities we interpolate between the stellar parameters of these isochrones whereas for [Fe/H] > 0.0 we (linearly) extrapolate the stellar parameters of the templates at the mass data points of the Lyon isochrones.

As in Conroy & van Dokkum (2012a), we ensure a monotonic increase in the mass along the isochrone by shifting the mass scale of the Padova isochrones such that they match the mass scale of the Dartmouth isochrones at the point where the two are joined together. For more details with respect to these stitched isochrones and the rationale behind choosing this particular combination we refer the reader to Conroy & van Dokkum (2012a).

The Parsec isochrones and the stitched isochrones that we use in the model are defined for the same age-metallicity grid. That age-metallicity grid ranges from $\log t = 9.175$ to $\log t = 10.125$ (i.e. 1.5 – 13.3 Gyr) with $\Delta \log t = 0.025$ and from [Fe/H] $= -1.4$ to [Fe/H] $= 0.4$ with Δ[Fe/H] $= 0.05$. Figure 4.4 shows the youngest and oldest isochrones at solar metallicity for both sets of isochrones.

As a final step, we combine the stellar parameters of the isochrones with the (optimized) MIX interpolator to create the stellar templates for the model. The interpolated spectra are rebinned to the wavelength grid of the SDSS data that we aim to fit in this paper.

4.4 The data

We apply our model to a set of stacked SDSS spectra of ETGs (see Section 4.5) selected from SDSS DR 12 (Alam et al. 2015) and similar to Spiniello et al. (2014). In the selection process we adopt the following selection criteria. We require that:

1. The equivalent width of Hβ and [O III]4959 is less than 0.5 to avoid strong emission lines.

2. The redshift is $z \leq 0.05$ so that both NaI and CaT are covered by all spectra.

3. The median SNR of the spectrum is more than 20 per angstrom (over the complete wavelength range) to increase the final SNR of the stack.

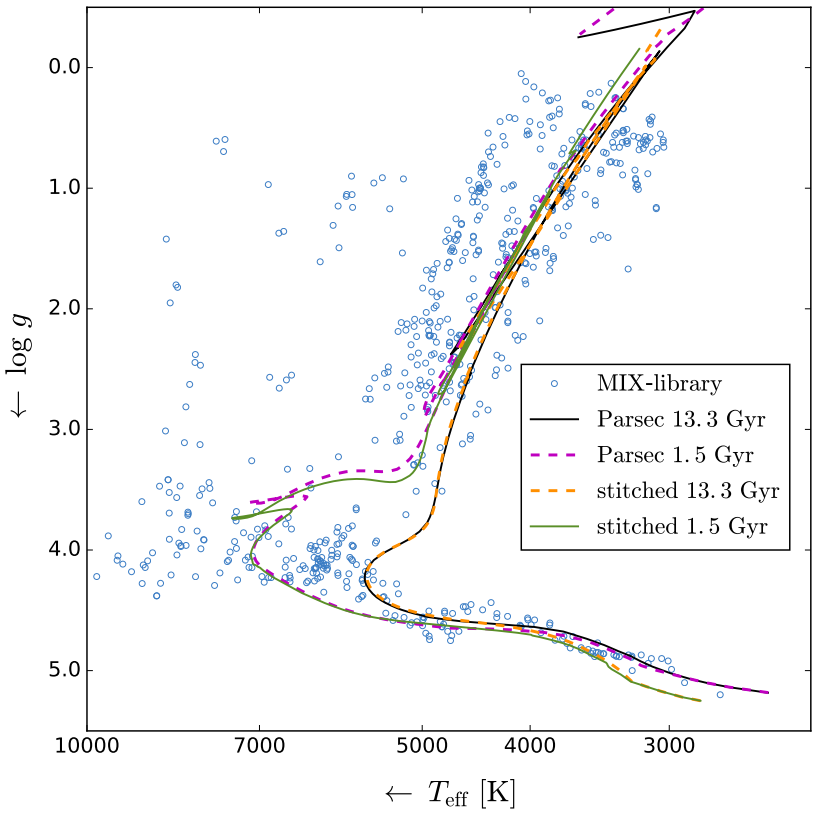

Figure 4.4 – Youngest and oldest isochrones in the adopted age-metallicity grid for the two sets of isochrones considered in this work. The *solid black* line corresponds to a Parsec isochrone with $t = 13.3$ Gyr, the *dashed magenta* line to a Parsec isochrone with $t = 1.5$ Gyr, the *dashed orange* line to a stitched isochrone with $t = 13.3$ Gyr and the *solid green* line to a stitched isochrone with $t = 1.5$ Gyr. The *blue circles* correspond to the stars in the MIX library. All isochrones in this figure correspond to solar metallicity.

4. The weight of the de Vaucouleurs' component in the composite fit of the surface brightness profile is equal to one in the SDSS g and r bands. This ensures the selection of ETGs.

At the selection stage, the galaxies are divided into six different velocity dispersion bins which are summarized in Table 4.2.

After the selection of the spectra, we stack the spectra in each of the velocity dispersion bins. In order to do this, we first shift the galaxy spectra

Table 4.2 – Overview of stacked SDSS spectra in six different velocity dispersion bins. The minimum and maximum velocity dispersion of each bin are indicated by $\sigma_{\rm min}$ and $\sigma_{\rm max}$ whereas $n_{\rm gal}$ represent the number of galaxies in each bin. $\mathrm{SNR}_{\rm stack}$ corresponds to the median SNR of the stacked spectrum and is calculated as the median of the flux divided by the median of the error over the complete wavelength range.

name	$\sigma_{\rm min}$ [km s^{-1}]	$\sigma_{\rm max}$ [km s^{-1}]	$n_{\rm gal}$	$\mathrm{SNR}_{\rm stack}$
SDSS170	140	170	1004	1128
SDSS200	171	200	737	1054
SDSS230	201	230	439	889
SDSS260	231	260	231	687
SDSS290	261	290	81	410
SDSS320	291	320	19	206

to the rest frame and normalize the spectra in the wavelength range $\lambda = 5400-5600\,\text{Å}$. Then the spectrum of each galaxy is smoothed to the highest velocity dispersion of the corresponding bin so that all spectra in a given bin have the same velocity dispersion. To create the stacked spectrum $\boldsymbol{s}_{\rm stack}$, every spectrum in a bin is assigned a weight

$$\boldsymbol{w} = \frac{1}{\mathbf{var}}, \tag{4.7}$$

with **var** the variance. Note that the variance is a vector so that different bins can have different weights. The stacked spectrum is then calculated as

$$\boldsymbol{s}_{\rm stack} = \frac{\sum\limits_i \boldsymbol{w}_i \boldsymbol{s}_i}{\sum\limits_i \boldsymbol{w}_i}, \tag{4.8}$$

and the corresponding error spectrum $\boldsymbol{e}_{\rm stack}$ as

$$\boldsymbol{e}_{\rm stack} = \frac{1}{\sqrt{\sum\limits_i \boldsymbol{w}_i}}. \tag{4.9}$$

Table 4.2 provides an overview of the six stacked SDSS spectra resulting from the stacking procedure, the number of galaxies that was used in each of these bins and their stacked SNRs. Figure 4.5 shows the six stacked SDSS spectra smoothed to the same velocity dispersion of 350 km/s and the difference between the SDSS spectra and the average spectrum of the stacked spectra.

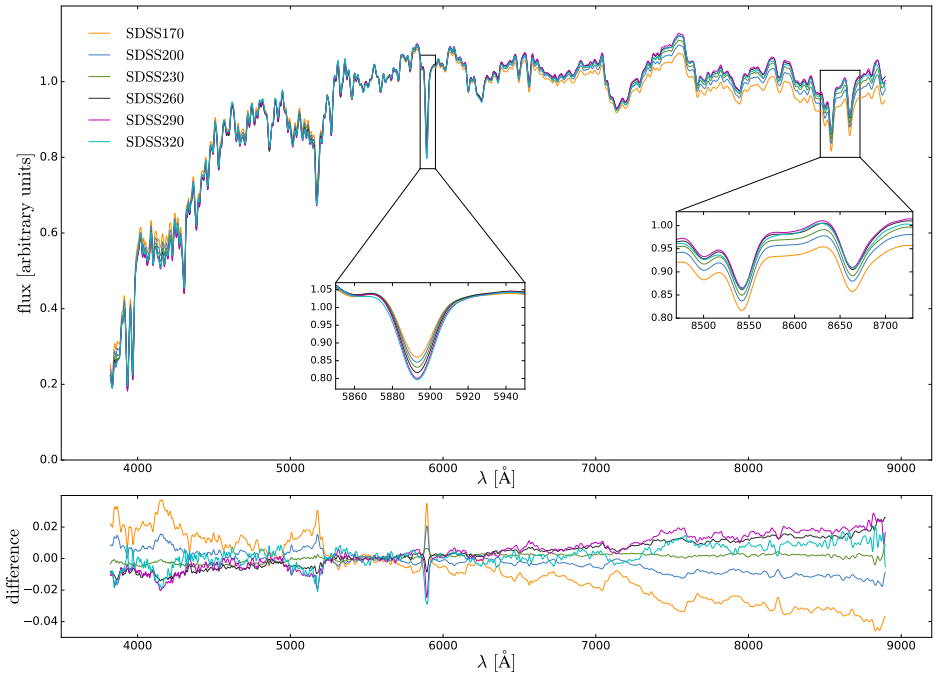

Figure 4.5 – Overview of the six stacked SDSS spectra considered in this work. The top panel shows the spectra that are binned as a function of velocity dispersion. Also shown are two zoom-ins, the first zoom-in shows the NaD-region and the second zoom-in the calcium triplet. All the spectra are smoothed to 350 km/s. The bottom panel shows the difference between the six SDSS spectra and the average spectrum of these six spectra.

Table 4.3 – Options for the different model ingredients.

model ingredient	options
isochrones	stitched, Parsec
regularization scheme	identity, $1/\boldsymbol{w}_0^2$
IMF prior parameterization	SPL, DPL, DPL with prior
abundance variations	Mg, Ca, Si, Ti, Na

4.5 Results

In this section we apply our model with a variety of different ingredients to the stacked SDSS spectra discussed in Section 4.4. For all the different runs that we carry out, we include a tenth-order multiplicative polynomial in the fit of both the parameterized version and the full version of the code (see Section 4.2). To account for the redshift range of our galaxy sample and to ensure that all wavelength bins have a sufficient number of galaxies, we cut all spectra above 8800 Å.

We use local Gaussian covariance structures at the emission lines [O III]4959, [O III]5007, [N II]6548 and [N II]6584. The height of these covariance structures is chosen such that it represents an additional error that corresponds to 3% of the average value of the spectrum, and the standard deviation of the Gaussian structures is 2.5 Å. These values are chosen visually to provide a good mask for the emission lines in the fits. Furthermore, we add boxcar covariance structures in the regions 6860-6920 Å, 7160-7360 Å, 7580-7680 Å and 8130-8360 Å. These regions are heavily contaminated by telluric absorption lines. The boxcar covariance structures increase the rms in the telluric regions by a factor of three based on our estimate of the uncertainties introduced by the spectral corrections.

We consider different model ingredients that we objectively compare on the basis of the Bayesian evidence. As discussed in Section 4.3.5, we consider two sets of stellar templates that are based on different isochrones. When we apply the model to the data we also consider two different regularization schemes. One regularization scheme in which $\mathbf{C}_{\rm pr}^{-1} = \mathbf{I}$, i.e. the identity matrix, and one in which

$$\mathbf{C}_{{\rm pr},ii}^{-1} = \frac{1}{\boldsymbol{w}_{0,i}{}^2} \tag{4.10}$$

with $\mathbf{C}^{-1}_{\mathrm{pr},ii}$ a diagonal matrix. We consider three different parameterizations of the IMF prior: a single power law (SPL), a double power law (DPL) with a break at 0.5 $\mathrm{M}_{\odot}$ and the same DPWL with an additional prior on α_2 such that

$$\log \mathrm{p}(\alpha_2) = -\frac{1}{2}\frac{(\alpha_2 - 2.35)^2}{0.15^2} - \log(2\pi \cdot 0.15^2). \tag{4.11}$$

This additional prior allows us to constrain the degeneracy between the two IMF slopes (see DTK16). Finally, we consider abundance variations of the elements Mg, Ca, Si, Ti and Na using the response functions introduced in Section 4.2.2.1. To reduce the already considerable computation time on a 196-CPU cluster, we do not consider response functions of more than two elements at the same time. The different model ingredients that we consider are summarized in Table 4.3.

When we apply the model to the SDSS data we limit the number of SSPs to $n_{\mathrm{SSPs}} \leq 4$, since the computational time scales increase significantly when we include an additional SSP. Note that according to DTKPS18, $N = 2-3$ SSPs should be sufficient to remove most of the bias on the inferred IMF, so we expect that using $N_{\mathrm{max}} = 4$ is sufficient here. We then select the number of SSPs that provides the highest evidence for every spectrum in all the runs and we use the corresponding results in the rest of our analysis.

In total, we carry out twelve different runs with different combinations of the model ingredients given in Table 4.3. The different runs that we consider are given in Table 4.4. We first consider the results of `run1` and `run2` in detail in Section 4.5.1. In Sections 4.5.2-4.5.5 we investigate the effect of including several response functions and changing the isochrones, the regularization scheme and the IMF parameterization. We assess different models in term of the Bayesian evidence.

4.5.1 Results `run1` and `run2`

The results of the first two runs are summarized in Table 4.5. The only difference between these two runs is that `run2` includes the response functions of Mg and that `run1` does not. If Mg is not included in the fit, the inferred IMF slopes are slightly biased to a higher value. According to the results in Table 4.5, the difference in log evidence between `run1` and `run2` is at least more than 100 in favour of `run2`. This implies that there is decisive evidence for including the response function of Mg in the fit (Jeffreys 1961). Therefore, we include the Mg response functions in all the fits after `run2`.

Table 4.4 – Overview of the different model runs and the ingredients that we use in each of these runs.

name	isochrones	reg. scheme	IMF prior parameterization	abundance variations
`run1`	stitched	identity	SPL	-
`run2`	**stitched**	**identity**	**SPL**	**Mg**
`run3`	stitched	identity	SPL	Mg, Ca
`run4`	stitched	identity	SPL	Mg, Si
`run5`	stitched	identity	SPL	Mg, Ti
`run6`	stitched	identity	SPL	Mg, Na
`run7`	Parsec	identity	SPL	Mg
`run8`	stitched	$1/\boldsymbol{w}_0^2$	SPL	Mg
`run9`	stitched	identity	DPL	Mg
`run10`	stitched	identity	DPL with prior	Mg
`run11`	stitched	identity	DPL with prior	Mg, Ti
`run12`	stitched	identity	DPL with prior	Mg, Na

Table 4.5 – Results for `run1` and `run2`. The number of SSPs that optimizes the evidence is indicated with $n_{\rm SSPs}$, the IMF slope with α and the strength of the Mg response functions with [Mg/Fe].

name	run	$n_{\rm SSPs}$	α	[Mg/Fe]	evidence
`SDSS170`	1	4	$2.19^{+0.02}_{-0.02}$	-	12775.4
`SDSS200`	1	3	$2.34^{+0.03}_{-0.03}$	-	12318.8
`SDSS230`	1	3	$2.46^{+0.02}_{-0.02}$	-	11939.5
`SDSS260`	1	3	$2.54^{+0.02}_{-0.02}$	-	11562.5
`SDSS290`	1	3	$2.65^{+0.02}_{-0.02}$	-	11131.9
`SDSS320`	1	3	$2.64^{+0.02}_{-0.02}$	-	10778.4
`SDSS170`	2	4	$2.07^{+0.03}_{-0.03}$	0.07	12956.7
`SDSS200`	2	3	$2.19^{+0.02}_{-0.02}$	0.10	12626.3
`SDSS230`	2	3	$2.33^{+0.02}_{-0.02}$	0.13	12331.7
`SDSS260`	2	3	$2.51^{+0.02}_{-0.02}$	0.15	12008.7
`SDSS290`	2	3	$2.44^{+0.03}_{-0.03}$	0.17	11579.0
`SDSS320`	2	2	$2.56^{+0.02}_{-0.02}$	0.18	11181.4

We will now consider in more detail the reconstructed ages and metallicities of the SSPs used in the fits, the reconstructed IMF and the reconstructed spectra for `run2`. Since the evidence for `run2` is much higher than for `run1`, we will use the results of `run2` as the reference model against which we compare the results of the other runs.

4.5.1.1 SSP ages and metallicities

In Figure 4.6 we show the mass- and luminosity-weighted ages and metallicities resulting from the fits to the SDSS spectra. Since the luminosity-weighted parameters are more affected by the younger components than the mass-weighted parameters, the luminosity weighted age is younger than the mass-weighted age and the luminosity-weighted metallicity is lower than the the mass-weighted metallicity. Both the luminosity-weighted and the mass-weighted ages and metallicities increase (slightly) as a function of velocity dispersion. This is in line with previous studies of ETGs (e.g., Conroy et al. 2014). However, the absolute values of the luminosity- and mass-weighted ages and metallicities are relatively high compared to previous studies, whereas the increase in age that we find is relatively small. Most likely this is related to using multiple SSPs in our fits: the younger components of the spectra are modelled by younger SSP components of the model and therefore the older components of the model are not biased as much by the younger components as they would have been if we had used only one SSP.

For most of the fits in `run1` and `run2` the model prefers fits with three SSPs. In Figure 4.7 we show the ages and metallicities of these SSPs for `run2`. For all the fits, the main component is an old metal-rich SSP that contributes between ∼65% and ∼80% to the integrated luminosity of the stacked SDSS spectra in the fit. The second component in the four lowest velocity dispersion bins is a young SSP ($t \sim 4$ Gyr) that is relatively metal-poor ([Fe/H] ~ -0.6). For `SDSS170` there is also an old metal-poor component ($t \sim 13\,\mathrm{Gyr}$, [Fe/H] ~ -0.6) that is not inferred from the other spectra. If we go from the lowest velocity dispersion bin to the highest velocity dispersion bin, the relative contribution of the old metal-rich component increases, while the relative contribution of the younger components decreases.

Although the ages and metallicities of the SSPs used in the fits should be related to the actual SFH of the galaxies in the stacked spectra, we note that the actual SFH of these objects is in reality extended. In that respect,

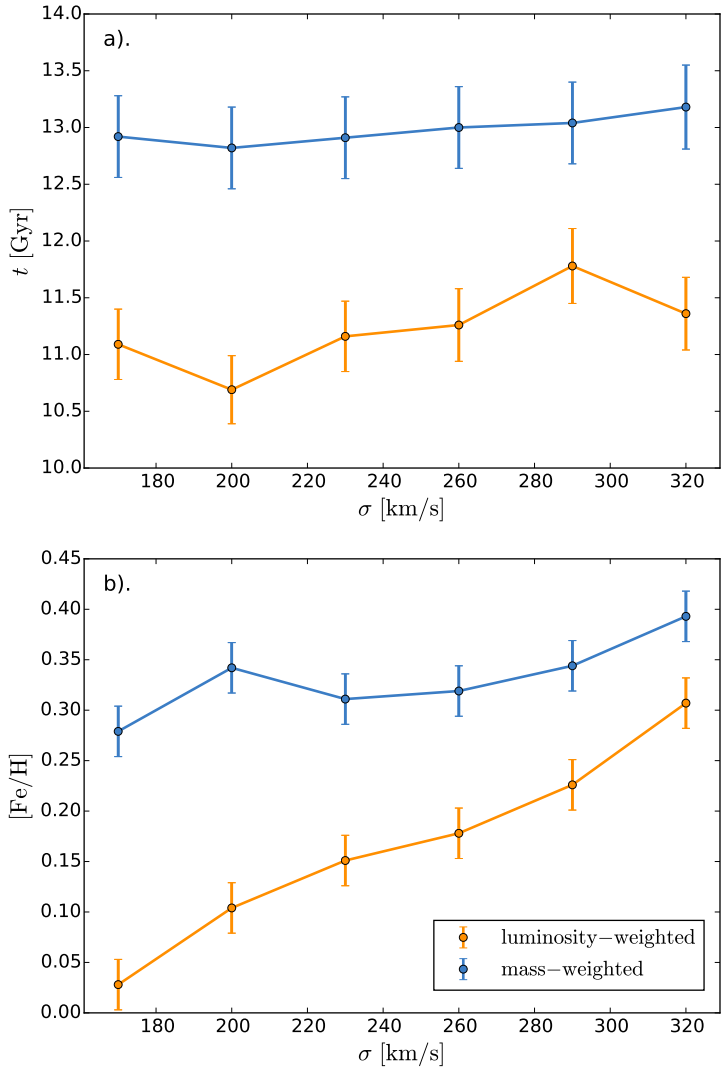

Figure 4.6 – Luminosity- and mass-weighted ages (panel *a*) and metallicities (panel *b*) for `run2` as a function of velocity dispersion for the stacked SDSS spectra. Errors are based on the spacings of the age-metallicity grid of the stellar templates.

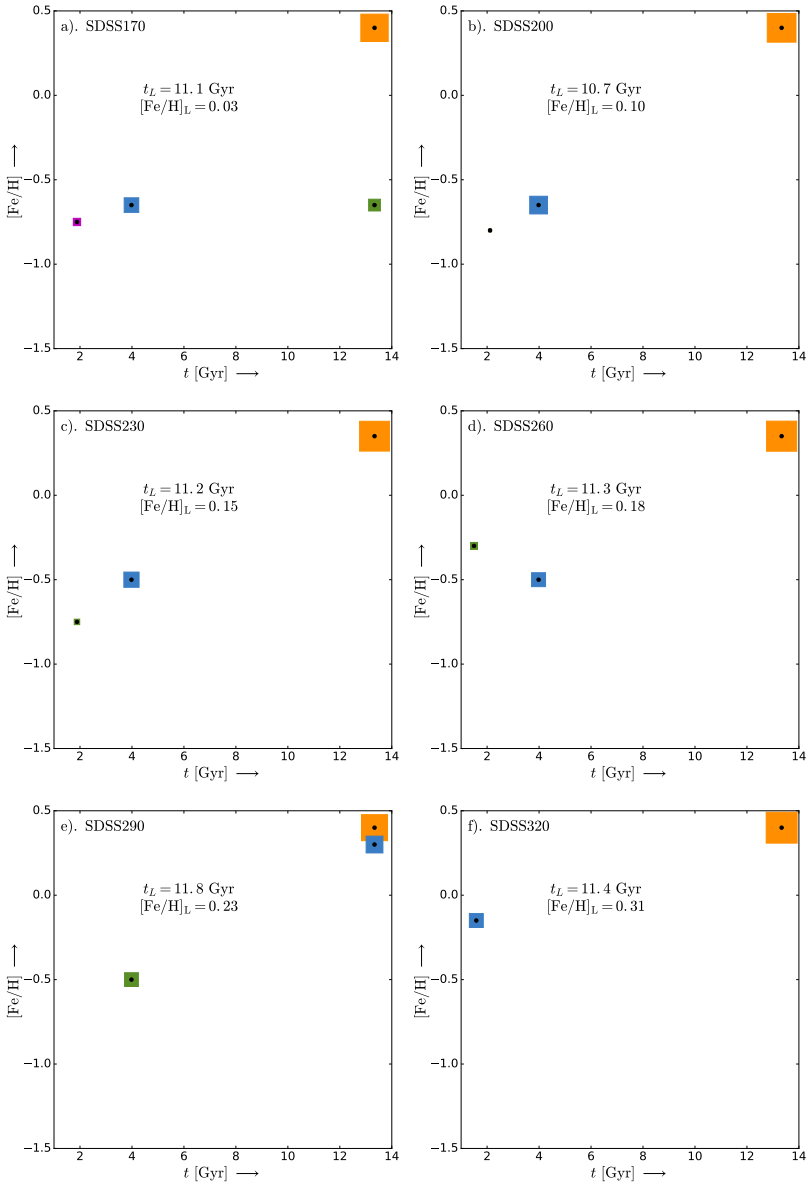

Figure 4.7 – Ages and metallicities of the SSPs used in `run2` for the most probable number of SSPs (four SSPs for `SDSS170`, two SSPs for `SDSS320` and three SSPs for the other spectra). The area of the squares representing the different SSPs corresponds to the relative contribution of that SSP to the total integrated luminosity. Also shown are the luminosity-weighted age t_L and metallicity $[Fe/H]_L$.

these ages and metallicities only allow us to reconstruct a coarse SFH, and with three or four SSPs, there may be degeneracies between the ages and metallicities of the different SSPs due to the age-metallicity degeneracy. Nevertheless, the superposition of the SSPs used by the model in the fits changes very little for the different runs and shows evidence for only a small amount of more recent star formation in these galaxies. As shown in DTKPS18, using multiple SSPs is a far better representation of such data than a single SSP and significantly reduces biases in the IMF inference.

The relative contribution of the younger components varies between $\sim$17% and $\sim$27% in terms of luminosity. However, in terms of stellar mass the contribution of this population is only between $\sim$1% and $\sim$7%. Therefore, a relatively small amount of recent star formation in terms of mass, for example due to the accretion of a low-mass satellite, may boost the luminosity of the galaxy significantly.

4.5.1.2 IMF reconstructions

In Figure 4.8 we show the reconstructed IMF of `SDSS230` with three SSPs for `run2`. The IMF and the prior IMF are very similar in this case. Although there appears to be a significant deviation for stars around 1.5 $M_{\odot}$, the number of stars in that particular bin is actually very small. This is both because the contribution of that SSP to the spectrum is small and because the mass bin is very narrow. This implies that a relatively small deviation from the prior on the weights in terms of the absolute number of stars may result in a large, but not significant, deviation of the IMF, since the IMF is obtained by dividing by the width of the mass bin, which is amplified as a consequence of the log-scale of the plot.

The fact that the reconstructed IMF and the prior IMF in Figure 4.8 are very similar is related to the regularization scheme that we use. Using the identity matrix as a regularization scheme penalizes the absolute deviation of the weights from the prior on the weights. The highest mass bins correspond to the most luminous stellar templates. Therefore deviating from the prior on the weights is most effective for these high mass bins, which is why with this regularization scheme deviations between the reconstructed IMF and the prior IMF are usually found in the highest mass bins. However, if we use multiple SSPs in the fit, these deviations are relatively small for this regularization scheme. The advantage of the great similarity between the reconstructed IMF and the prior IMF is that we can

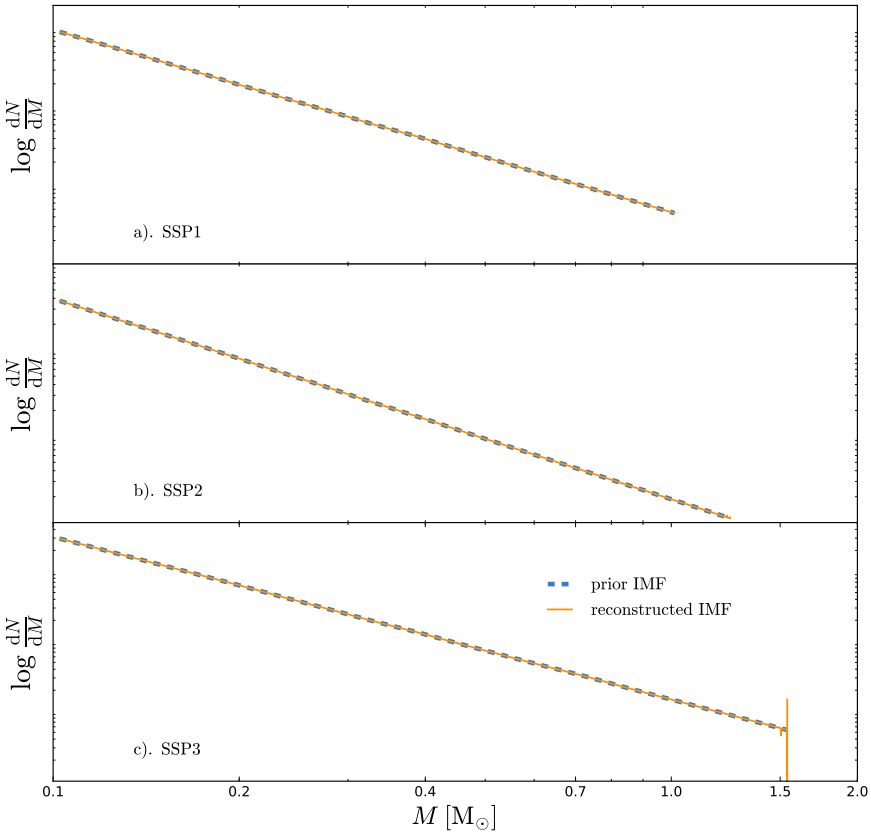

Figure 4.8 – Reconstructed IMF of `SDSS230` in `run2`. The *orange* lines correspond to the reconstructed IMFs and the *blue-dashed* lines to the prior IMFs. The different panels *a-c* correspond to the different SSPs used in the fit.

easily summarize our results in terms of the IMF prior parameters, e.g. one or two IMF slopes, which allows us to compare our results more easily to other studies. Another advantage is that this IMF prior results in realistic mass distributions that are relatively smooth. This is not necessarily the case for all regularization schemes. The disadvantage of this regularization scheme is that it gives less flexibility to the model. We will mostly use the identity matrix as a regularization scheme, but in Section 4.5.4 we investigate the effect of using another regularization scheme.

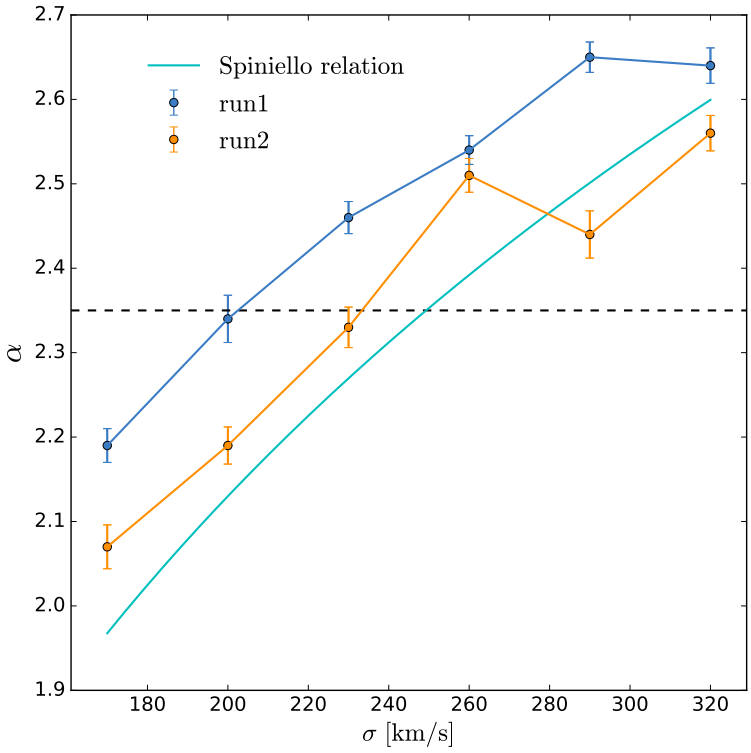

Figure 4.9 – Relation between velocity dispersion and IMF slope for `run1` (*blue* data points) and `run2` (*orange* data points). The *cyan* line represents the IMF slope–velocity dispersion relation of Spiniello et al. (2014) and the *black-dashed* line corresponds to a Salpeter IMF slope.

In Figure 4.9 we show the reconstructed IMF slopes as a function of velocity dispersion for `run1` and `run2` together with the IMF slope–velocity dispersion relation found by Spiniello et al. (2014) using an adapted code by Conroy & van Dokkum (2012a). If we do not use the response functions of Mg, the reconstructed IMF slopes appear to be biased to a value that is too high. Besides that, the trends that we recover for `run1` and `run2` are very similar. If we compare the results of `run1` and `run2` with the Spiniello et al. (2014) relation obtained using the code of Conroy & van Dokkum (2012a), we can conclude that the recovered trends are very similar although the result that we obtain in `run2` for `SDSS260` appears to be slighly high

compared to the other results. In an absolute sense, `run2` agrees better with the Spiniello relation, which we can understand from the fact that Spiniello et al. (2014) also take into account abundance ratio variations in the form of an [α/Fe]-parameter. Altogether, using a completely independent code based on full-spectrum fitting, the results in Figure 4.9 confirm earlier results in the literature, such as those by Spiniello et al. (2014) based on the analysis of line-strength indices, that for ETGs a relation exists between their stellar velocity dispersion and the IMF slopes.

4.5.1.3 Reconstructed spectra

As a final step in the analysis of `run1` and `run2` we consider the reconstructed spectra. In Figure 4.10 we show the reconstructed spectrum of `SDSS230` for `run2` with three SSPs. Overall, the original spectrum and the reconstructed spectrum in the top panel match well. However, we learn more by considering the residual between the original and the reconstructed spectrum in the lower three panels of Figure 4.10.

The shaded-orange regions in the residual-panels correspond to the error spectrum adopted by the model. This error spectrum includes the original error spectrum of the data, the local covariance structures for the emission lines and the strong telluric regions and an additional global covariance b to account for systematic uncertainties. As in DTKPS18, the extra global covariance b is parameterized with the parameter $b_{\rm cov}$ such that

$$b = b_{\rm cov} \cdot \text{median}\left(\mathbf{C}_{\rm D,old}\right), \tag{4.12}$$

where the original covariance matrix $\mathbf{C}_{\rm D,old}$ and the new covariance matrix $\mathbf{C}_{\rm D,new}$ relate as

$$\mathbf{C}_{\rm D,new} = \mathbf{C}_{\rm D,old} + b\mathbf{I}. \tag{4.13}$$

The SNR of the stacked spectra is very high and varies between $\sim$200 and $\sim$1100 (see Table 4.2). Since the model is not able to fit the data to that level of accuracy, it increases the value of $b_{\rm cov}$. In all our fits the systematic uncertainties parameterized by $b_{\rm cov}$ are significantly higher than the intrinsic uncertainties of the spectra. For `SDSS320` the systematic uncertainties are $\sim$1.5 times higher than the intrinsic uncertainties whereas for `SDSS170` the systematic uncertainties are $\sim$7 times higher than the intrinsic uncertainties.

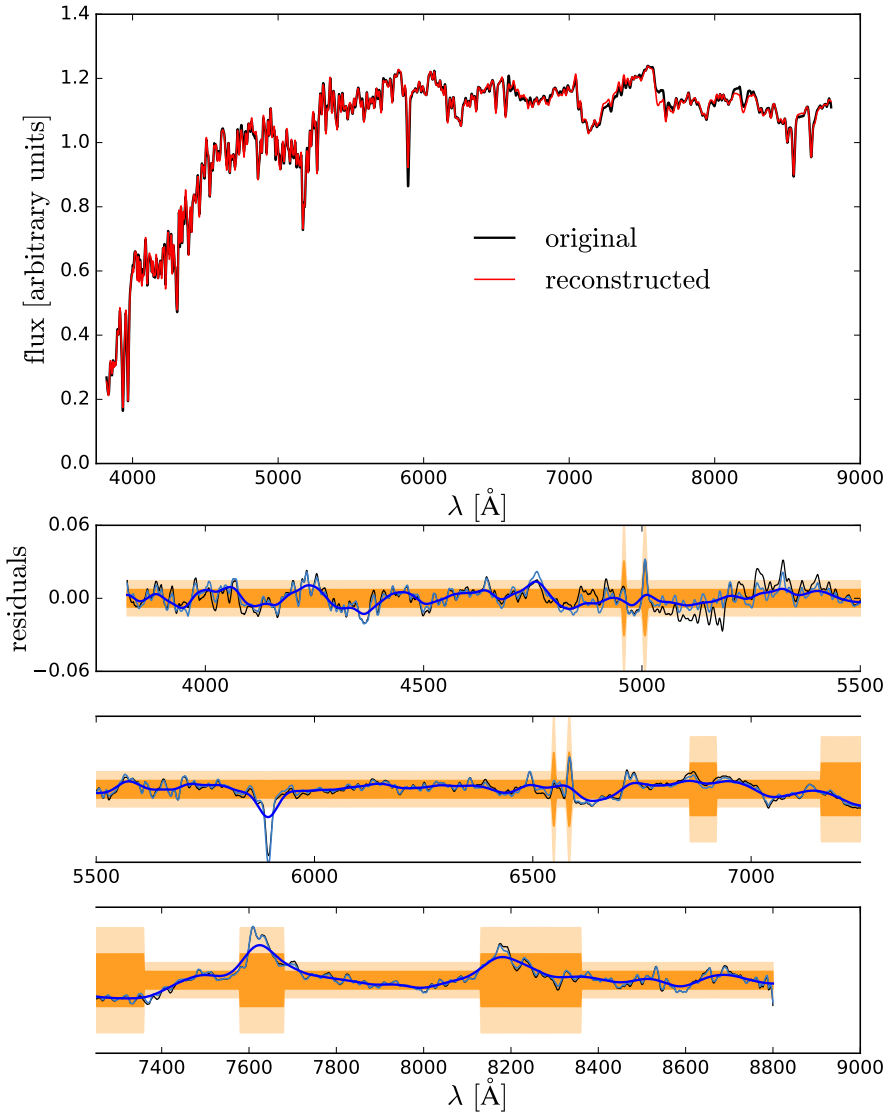

Figure 4.10 – Reconstructed spectrum of `SDSS230` obtained in `run2` with three SSPs. The upper panel shows the original spectrum (*black*) and the reconstructed spectrum (*red*). The three bottom panels show the residual between the original spectrum and the reconstructed spectrum (*thin-blue* line). Also shown in the residual panels are a smoothed version of the residual (*thick-blue* line), the error spectrum adopted by the model (*shaded-orange* region), and the residual obtained in `run1` for the same spectrum (*thin-black* line).

Taking into account the different sources of uncertainty in the covariance matrix, the residuals (thin-blue line) in Figure 4.10 are in general well within the error spectrum adopted by the model and represented by the shaded-orange regions. An exception to this is the sodium doublet NaD at $\sim$5890 Å, which does not fit well. The reconstructed spectra for the other stacks look similar and also do not fit NaD correctly. This is consistent with Spiniello et al. (2015) who also find that NaD is a very troublesome line. The thick-blue line in the residual-panels corresponds to a smoothed version of the residual. This line shows clearly that there are strong deviations between the model and the original spectrum in the telluric regions. However, these deviations are expected and we account for this in the covariance matrix. At the blue end of the spectrum, the fit shows an oscillatory pattern. We suspect that this is at least partly related to the small flux level in these regions but there might also be issues with the flux calibration or the response functions.

The thin-black line in the residual-panels represents the residuals of the most probable fit of `SDSS230` obtained in `run1`. If one carefully looks at this line, it becomes clear that the Mg response functions help to resolve some important issues in the region 5000-5500 Å where many iron and magnesium lines are found.

4.5.2 Abundance variations

As a next step we investigate the effect of including the response functions of Ca, Si, Ti and Na on the inferred IMF slope–velocity dispersion relation. The reason for testing Ca, Si and Ti is that these elements are also alpha-elements (like Mg), and ETGs are known to be alpha-enhanced with respect to the Milky Way due to their different SFHs (e.g. Trager et al. 2000a; Johansson et al. 2012; Conroy et al. 2014). We test the inclusion of the response functions of Na because the model was not able to fit the NaD-line correctly in `run2`.

In Figure 4.11 we show the IMF slope–velocity dispersion relation that we recover for `run3-run6` by including the response functions of Ca, Si, Ti and Na. For all these runs we include the response functions of Mg as well. We also show the results of `run2` in Figure 4.11 to clearly show the effect of including the various response functions.

Including the response functions of Ca and Si has almost no effect on the inferred IMF slope–velocity dispersion relation, except for the data points of

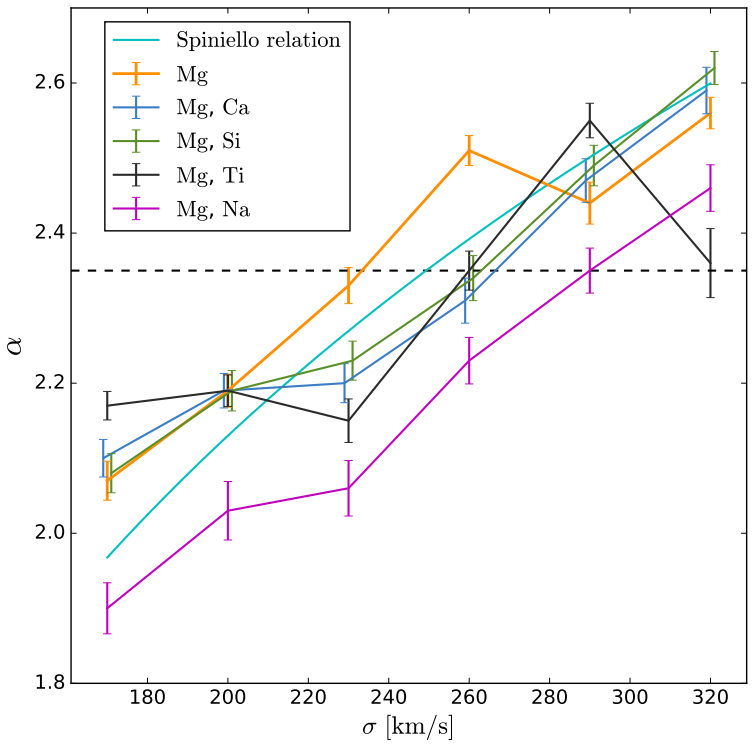

Figure 4.11 – IMF slope–velocity dispersion relations obtained by including various response functions in the model. The *cyan* line corresponds to the IMF slope–velocity dispersion relation of Spiniello et al. (2014) and the *black-dashed* line corresponds to a Salpeter IMF slope.

SDSS230 and SDSS260, which are slightly lower when Ca or Si are included. When the response functions of Ti are included, we still recover a similar IMF slope–velocity dispersion relation, but this time the relation is more irregular. In particular the last data point of SDSS320 is $\Delta\alpha \sim 0.3$ lower when Ti is included. Finally, if we include the response functions of Na, the trend between IMF slope and velocity dispersion remains very similar but is shifted lower by $\Delta\alpha \sim 0.2$. To investigate this further, in Figure 4.12 we show a zoom-in of the fit of the sodium doublet that we obtain with and without the Na response functions. This figure clearly demonstrates that by including the Na response functions improves the fit of the NaD region

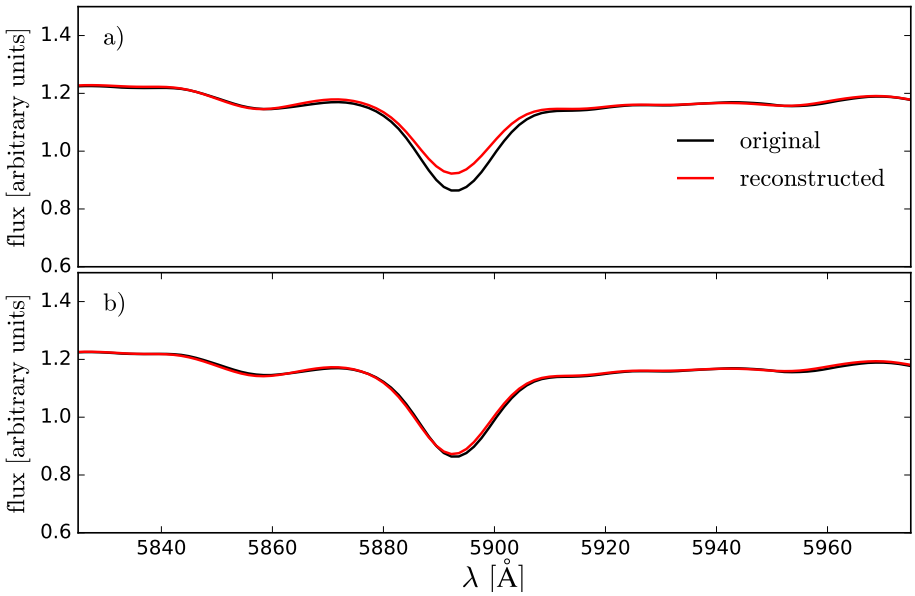

Figure 4.12 – Zoom-in of the original spectrum (*black*) and the reconstructed spectrum (*red*) in the wavelength region of the sodium (Na) doublet. Panel *a* is without the Na response functions and panel *b* with the Na response functions.

and confirms that the NaD doublet is very sensitive to [Na/Fe] variations (e.g. Spiniello et al. 2015)

The abundance variations that we find for the different elements as a function of velocity dispersion are shown in Figure 4.13. As expected, the abundance of the alpha-elements Mg and Si increases as a function of velocity dispersion. The abundance of Ca is almost constant as a function of velocity dispersion and is close to zero. This result confirms earlier results by, e.g., Worthey (1998), Trager et al. (2000b) and Thomas et al. (2003) who find that Ca tracks Fe and that Ca as an alpha-element is underabundant with respect to Mg. For Ti there appears to be a small anti-correlation between the velocity dispersion and [Ti/Fe] which is not consistent with the results obtained by for example Johansson et al. (2012) and Conroy et al. (2014).

In Table 4.6 we show the evidence that we obtain after applying our model to the stacked spectra in `run2-run8`. This table shows that including the response functions of Na has the highest impact on the evidence and that there is decisive evidence for including the Na response functions for

Table 4.6 – Evidence obtained by including different combinations of response functions (`run2`–`run6`), using different isochrones (`run2` vs. `run7`) and using different regularization schemes (`run2` vs. `run8`). The evidence in this table is the value that we find for the most probable number of SSPs. In the second column, S refers to the stitched isochrones and P to the Parsec isochrones.

name	iso.	reg. sch.	IMF prior	abundance variations	ev. SDSS170	ev. SDSS200	ev. SDSS230	ev. SDSS260	ev. SDSS290	ev. SDSS320
run2	S	I	SPL	Mg	12956.6	12626.3	12321.1	12008.7	11579.0	11181.4
run3	S	I	SPL	Mg, Ca	12960.1	12628.5	12328.9	12022.8	11594.3	11237.8
run4	S	I	SPL	Mg, Si	12953.1	12626.2	12332.5	12037.5	11618.6	11204.7
run5	S	I	SPL	Mg, Ti	12956.8	12623.9	12350.2	12037.6	11633.7	11268.2
run6	S	I	SPL	Mg, Na	13099.1	12887.8	12721.6	12480.6	12102.6	11809.3
run7	P	I	SPL	Mg	12884.7	12497.7	12159.0	11849.8	11454.5	11071.0
run8	S	$1/\boldsymbol{w}_0^2$	SPL	Mg	13003.8	12660.0	12357.9	12077.6	11728.6	11310.9

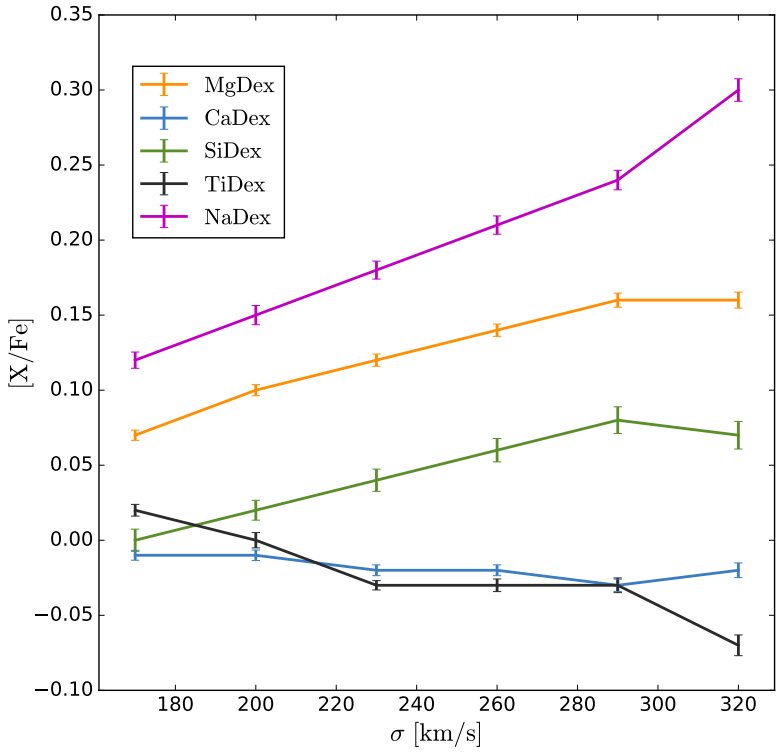

Figure 4.13 – Abundance variations as a function of velocity dispersion for the elements for which we consider the response functions. The vertical axis represents the abundance in [X/Fe] as defined in Section 4.2.2.1.

all spectra. When Ca is included, the effect on the evidence is relatively small except for `SDSS320`. For Si and Ti, the evidence for including the corresponding response functions increases with velocity dispersion, which is probably related to the increase/decrease of [Si/Fe]/[Ti/Fe] as a function of velocity dispersion. Apart from Mg and Na, we find that including Ti has the highest impact on the evidence. *Since we want to limit the total number of runs that we consider, in the remainder of this work we only consider the response functions of Mg, Ti and Na.*

All of the stacked spectra appear to be enhanced in Na and the value of [Na/Fe] increases as a function of velocity dispersion. This confirms earlier results by, e.g., Conroy & van Dokkum (2012b), Spiniello et al.

(2014) and La Barbera et al. (2017), who find that ETGs require sodium-enhanced population synthesis models. The reason for this enhancement is not completely clear yet and may be related to the IMF, a non-solar abundance pattern or the interstellar medium (ISM). Jeong et al. (2013) and Park et al. (2015) conclude that for ETGs the ISM is not an important factor and that the IMF alone cannot explain the sodium excess in these objects. Therefore ETGs should be sodium-enhanced with respect to the Milky Way.

One of the main concerns that we have regarding the use of the response functions of Ti and Na is that we make the assumption that these response functions are independent of the IMF. We only use response functions that are derived for a Salpeter IMF, interpolate this as a function of age and metallicity and then for every SSP apply the same response function to all stellar templates. However, if we compare the response functions derived for a Salpeter IMF with those derived for a Kroupa IMF we find that the response functions are not independent of the IMF. For Mg, the response functions for the two IMFs are relatively similar but for Ti and Na the differences are more significant. To improve the analysis of the effect of response functions even further, it would be better to apply individual stellar response functions to the spectra in matrix $\mathbf{S}$ instead of SSP-based response functions. Using individual response functions ensures that there are no degeneracies between the response functions and the inferred IMF parameters. However, these individual response functions are currently not available to us, which is why we choose to use the SSP-based response functions. In Appendix 4.A we investigate in more detail the consequences of using these response functions for the parameters that we infer.

4.5.3 Isochrones

To investigate the effect that the adopted set of isochrones might have on the inferred IMF slope–velocity dispersion relation, we have created a different set of stellar templates that is based on the Parsec isochrones. In `run7` we repeat `run2` using the stellar templates based on the Parsec isochrones. The comparison between the inferred IMF slope–velocity dispersion relation for `run2` and `run7` is shown in Figure 4.14. For $\sigma \leq 260\,\mathrm{km\,s^{-1}}$, the trends for both relations are similar. However, the relation is $\Delta\alpha \sim 0.2 - 0.3$ higher for the Parsec isochrones than for the stitched isochrones. In an absolute sense, the last two data points (`SDSS290`

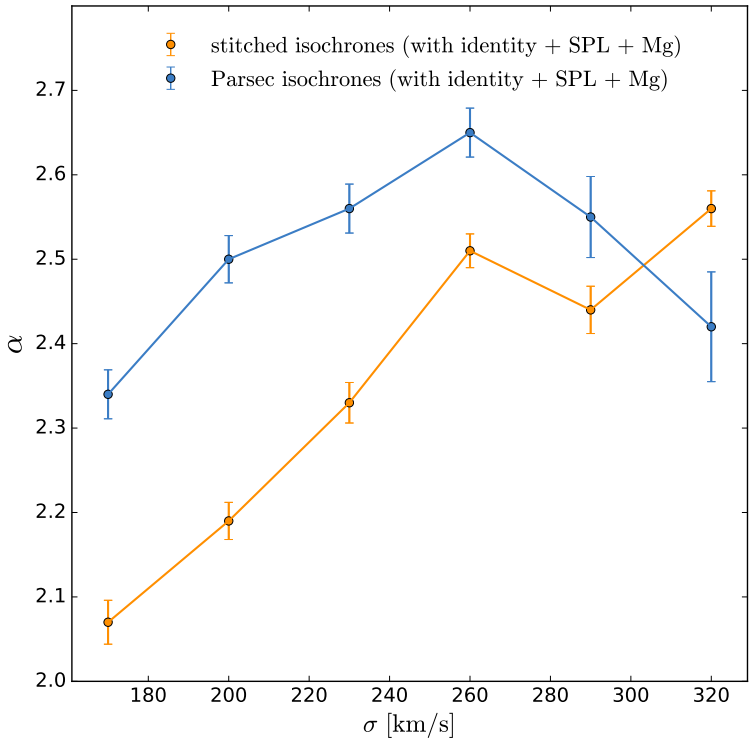

Figure 4.14 – IMF slope–velocity dispersion relation for the two sets of isochrones considered in `run2` and `run7`. For all the spectra there is very strong evidence in favour of the stitched isochrones (on average the Bayesian evidence is 124.7 higher for the stitched isochrones, i.e. a $\sim 15\,\sigma$ preference of these isochrones over the Parsec isochrones).

and `SDSS320`) of the two sets of isochrones are relatively close together. However, the trend for the two sets of isochrones is distinctly different, showing a decrease of the IMF slope after `SDSS260` for the Parsec isochrones.

There are a number of differences between the two sets of isochrones that might be the origin of the different IMF slope–velocity dispersion relations. First, the low-mass main-sequence tracks of the two isochrones sets are different, as can already be seen in Figure 4.4. Second, there may be small differences between the temperatures of the giant branches of the isochrone sets. These offsets may result in interpolated spectra that are slightly different. Finally, we use the V-band magnitudes of the isochrones to

normalize the interpolated spectra. These magnitudes may also be slightly different for the two isochrone sets. Altogether, small offsets between the isochrone tracks and a different normalization may result in a different set of normalized interpolated spectra. This may in turn affect the inference of the IMF.

The advantage of the Bayesian framework that our model provides is that we can compare the two sets of isochrones on the basis of the Bayesian evidence. In Table 4.6 we show the evidence that we obtain in `run2` and `run7` for the different spectra. This table clearly shows that for all spectra there is very strong to decisive evidence in favour of the stitched isochrones over the Parsec isochrones. Therefore, we will use the stellar templates based on the stitched isochrones in the remainder of this work.

4.5.4 The regularization scheme

In Figure 4.15 we show a comparison between the inferred IMF slope–velocity dispersion relation for `run2` and `run8`. The only difference between these two runs is the adopted regularization scheme as discussed at the beginning of Section 4.5. We conclude that the IMF slope–velocity dispersion relation is very similar for both regularization schemes.

Although the inferred IMF slope–velocity dispersion relation looks very similar for `run2` and `run8`, the actual reconstructed IMF for an individual spectrum may be quite different. In Figure 4.16 we show the reconstructed IMF of `SDSS230` in `run8` with four SSPs (four SSPs maximizes the evidence in this case). If we compare Figure 4.16 with Figure 4.8, it becomes clear that the $1/w_0^2$-regularization scheme of `run8` gives more freedom to the model and makes it easier to deviate from the prior.

In Table 4.6 we show the evidences that we obtain for `run2` and `run8`. For all spectra there is very strong to decisive evidence in favour of using the $1/w_0^2$-regularization scheme. Nevertheless, we use the identity matrix in almost all of the runs. The reason for this is that if we use the $1/w_0^2$-regularization scheme, the reconstructed IMF may sometimes become very irregular. Physically, there is no reason to expect such a very irregular IMF. In future work, it might be worthwhile to further investigate regularization schemes that focus more on the physical nature and smoothness of the IMF.

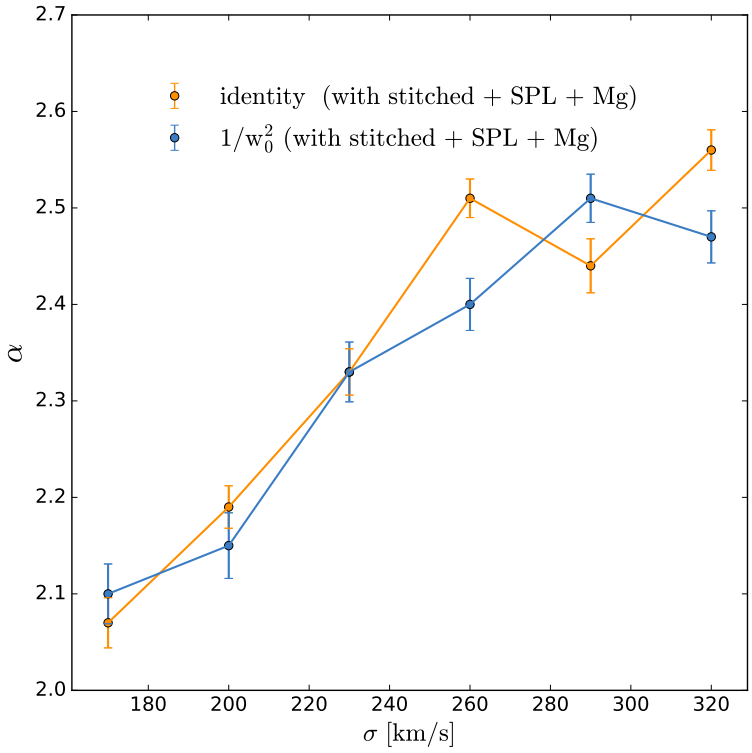

Figure 4.15 – IMF slope–velocity dispersion relation for the two regularization schemes considered in `run2` and `run8`.

4.5.5 The IMF parameterization

Various studies rule out a single power law IMF parameterization on the basis of constraints on the mass-to-light ratio or stellar population analysis (e.g., La Barbera et al. 2013, 2016; Lyubenova et al. 2016). Therefore, as a final step we investigate the effect of using a double power law parameterization of the IMF prior instead of a single power law parameterization. The double power law IMF parameterization that we use is characterized by a normalization, a low-mass slope α_1 and a high-mass slope α_2. The break between these two slopes is fixed at 0.5 $M_\odot$, chosen to be the same as the break between the low-mass IMF slope and high-mass IMF slope in the Milky Way (Kroupa 2001). However, in principle this

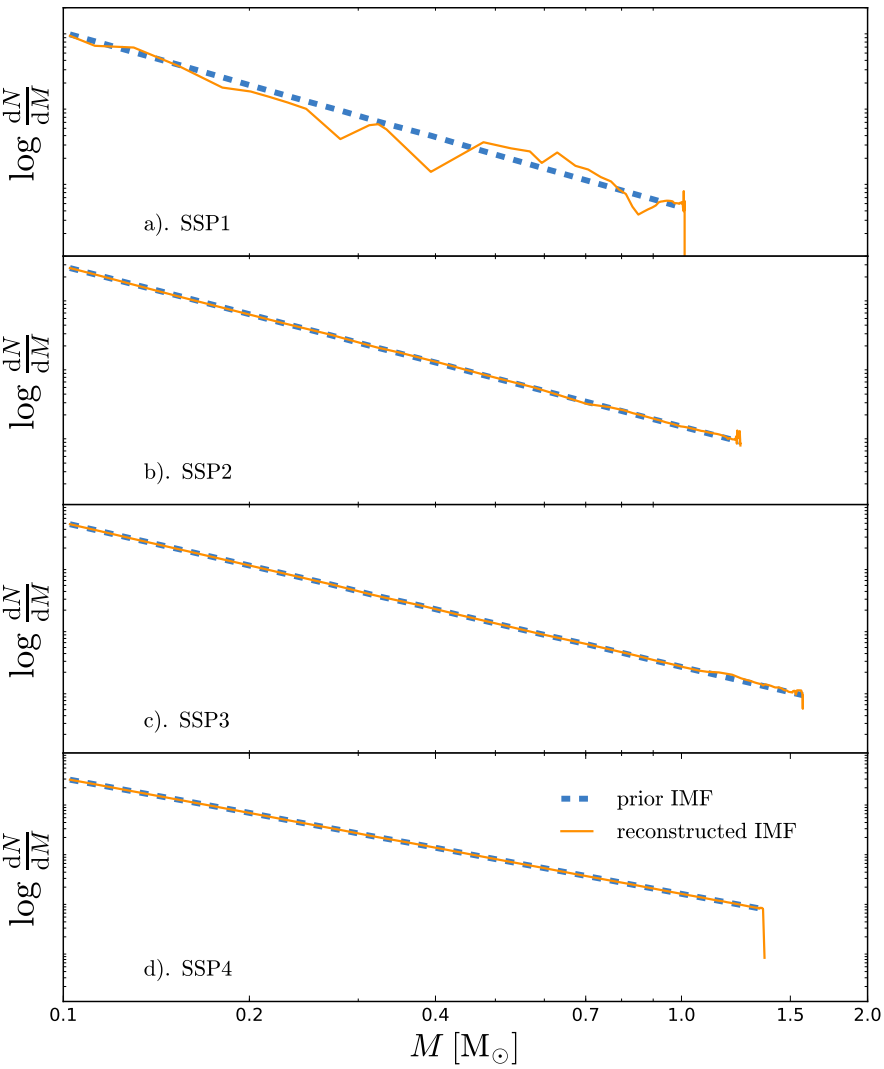

Figure 4.16 – Reconstructed IMF of `SDSS230` in `run8` with four SSPs. The *orange* lines correspond to the reconstructed IMFs and the *blue-dashed* lines to the prior IMFs. The different panels *a-d* correspond to the different SSPs used in the fit.

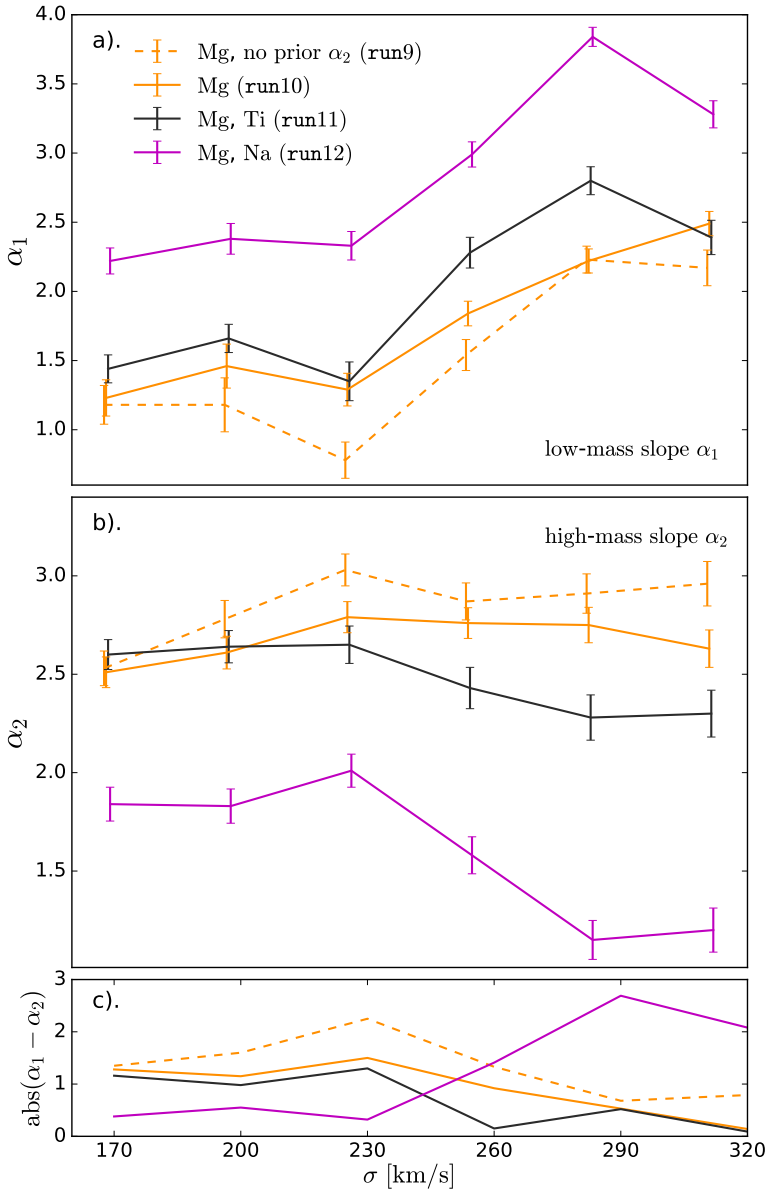

Figure 4.17 – Reconstructed IMF slopes as a function of velocity dispersion using a double power law parameterization of the IMF prior and including different combinations of response functions, corresponding to `run9`, `run10`, `run11` and `run12`. The reconstructed low-mass slopes α_1 are shown in panel *a* and the reconstructed high-mass slopes α_2 in panel *b*. Panel *c* shows the absolute difference between the two slopes.

could also have been a free parameter of the model. In `run9`, the two IMF slopes are completely free, whereas in `run10`, `run11` and `run12` we constrain the high-mass slope with an additional prior (see Equation 4.11).

The results of `run9`-`run12` are shown in Figure 4.17. For `run9` and `run10` we recover similar trends as a function of velocity dispersion showing an increase of the low-mass slope α_1 from approximately the value of a Kroupa IMF to the value of a Salpeter IMF, and a high-mass slope α_2 that is approximately constant but steeper than the Salpeter value. When we include the response functions of Ti, the trend for α_1 remains similar to the trends that are recovered when only Mg is included. However, the high-mass slope now shows a relatively small decrease as a function of velocity dispersion from slightly steeper than Salpeter to the Salpeter value. When the response function of Na are included, the trend that we find for the low-mass slope is similar to the previous trends but shifted to a significantly steeper value. At the same time, the high-mass slope appears to be much flatter and decreases as a function of velocity dispersion for $\sigma \geq 230\,\mathrm{km\,s^{-1}}$. However, the response functions of sodium are most likely not independent of the IMF (see also Appendix 4.A) and for a double power law IMF parameterization it becomes even harder to disentangle the effect of sodium enhancement from IMF variations. Therefore the results including sodium should be considered with caution.

In Table 4.7 we show the differences in log evidence between the runs with a double power law parameterization of the IMF and the corresponding runs with a single power law IMF parameterization. When using the Mg response functions, a double power law IMF is preferred in all cases over the single power law IMF. But this is not the case for the other runs, where there are a few cases for which a single power law is preferred. However, if one looks closely at those cases for which the evidence in favour of the double power law IMF is less than five, it becomes clear that for those cases the difference between the low-mass slope and the high-mass slope is rather small (see panel *c* of Figure 4.17). Therefore, it is to be expected that also a single power law IMF will provide a good fit to the data in those cases. Since the Bayesian framework of our model prefers models with fewer parameters, this may result in selecting a single power law IMF in favour of the double power law IMF. Given the strong preference for a double power law IMF parameterization in most of our results, and that in those cases where it is not preferred the IMF slopes are relatively close to

a single power law IMF, we conclude that ETGs should not be modelled with a single power law IMF, in particular for lower mass ETGs.

4.5.6 IMF shape from theoretical models of star formation

In Section 4.5.5 we have seen that there is strong evidence in favour of a double power law IMF parameterization with respect to a single power law IMF parameterization. Here, we investigate to what extent this result is compatible with theoretical models of star formation.

Stars form in dense molecular cores and therefore the IMF is expected to be related to the mass distribution of these cores, the core mass function (CMF) (e.g., Padoan & Nordlund 2002). One of the processes that is thought to play an important role in the process of star formation is turbulence. Turbulence is able to support molecular clouds against gravitational collapse on a global scale. However, dissipation of large-scale turbulent motions to smaller scales generates supersonic shocks and compressions which eventually may produce a filamentary structure with high density contrasts in which locally there are regions that are unstable against gravitational collapse (Chabrier 2003; Offner et al. 2014). Because the effects of turbulence are multiplicative, the central limit theory predicts that this picture of star formation naturally leads to a log-normal density distribution with a width that increases as a function of the Mach number $\mathcal{M}$ (Vazquez-Semadeni 1994; Padoan et al. 1997). The subsequent collapse into stars is determined by gravity.

Above some characteristic mass scale, most of the cores will collapse into stars as the masses of these cores are well above the typical Jeans mass of the star forming medium. Due to the turbulent fragmentation of cores and the resulting log-normal density distribution, some fraction of the cores with masses lower than the typical Jeans mass will have densities that are sufficiently high for gravitational collapse to occur and are therefore able to form stars. Although in reality the process of star formation is much more complicated than this simplified picture, assuming a simple mapping from the CMF to the IMF the natural outcome of this description is a Chabrier-like IMF with a log-normal distribution for the lowest stellar masses and a power law IMF for higher stellar masses (Chabrier 2003; Offner et al. 2014). More advanced descriptions using the (extended) formalism of Press & Schechter (1974) result in similar mass distributions (Hennebelle & Chabrier 2008; Hopkins 2012). The characteristic mass scale

Table 4.7 – Difference in log evidence when using a double power law parameterization of the IMF prior instead of a single power law parameterization of the IMF prior. The runs that are compared in this table all use the stitched isochrones and the identity matrix as regularization scheme.

	Δ log evidence			
name	`run2` (SPL, Mg) ↓ `run9` (DPL, Mg)	`run2` (SPL, Mg) ↓ `run10` (DPL+pr$_{\alpha_2}$, Mg)	`run5` (SPL, Mg, Ti) ↓ `run11` (DPL+pr$_{\alpha_2}$, Mg, Ti)	`run6` (SPL, Mg, Na) ↓ `run12` (DPL+pr$_{\alpha_2}$, Mg, Na)
SDSS170	5.7	6.0	12.2	17.2
SDSS200	15.6	11.9	11.3	-1.1
SDSS230	41.9	32.4	10.5	-2.3
SDSS260	16.1	13.0	-1.2	25.3
SDSS290	33.7	7.9	2.3	61.4
SDSS320	58.0	-2.1	-0.5	22.7

in this description is determined by the local properties of the star forming medium (i.e. $\mathcal{M}$) and may change for different typical environments of star formation.

The highest-mass ETGs are thought to form the bulk of their stellar mass at high redshift in a relatively short amount of time (e.g., Thomas et al. 2005; McDermid et al. 2015), which would imply that a significant fraction of their stellar mass forms in starbursts. In starburst galaxies, the typical conditions of star formation are expected to be different than in the Milky Way and characterized by more turbulence and therefore typically higher Mach numbers (Downes & Solomon 1998; Westmoquette et al. 2007; Greve et al. 2009). As a consequence of these higher Mach numbers, the characteristic mass of the initial mass function could be shifted to lower stellar masses and this would provide a possible explanation for the observed non-universal IMF in these objects.

Although from a theoretical point of view intrinsically different distributions, a double power law IMF and a log-normal distribution extended with a power law for higher masses can in fact result in very similar mass distributions (Dabringhausen et al. 2008). Therefore, when we fit the IMF with a double power law and find an increasing low-mass IMF slope, this may in reality be the consequence of a shift in the characteristic stellar mass of the log-normal distribution. This is illustrated schematically in Figure 4.18. The trends that we find of an increasing low-mass slope as a function of velocity dispersion and an almost constant high-mass slope are therefore consistent with a Chabrier-like IMF where the characteristic mass of the IMF shifts to lower stellar masses for higher velocity dispersions due to the different typical circumstances of star formation in higher velocity dispersion ETGs.

4.5.7 Mass-to-light ratios

Finally, in Figure 4.19 we show the r-band stellar mass-to-light ratios corresponding to the results of `run2`, `run8` and `run10`. To calculate these mass-to-light ratios we first determine the maximum-a-posteriori value of the sampled parameters and for these values we calculate the most probable weights. Then we multiply these most probable weights by the masses and r-band luminosities of the stellar templates to calculate the r-band mass-to-light ratio. We assume a low-mass cut-off of 0.1 $M_{\odot}$. The results in Figure 4.19 clearly show that the mass-to-light ratios that we find increase

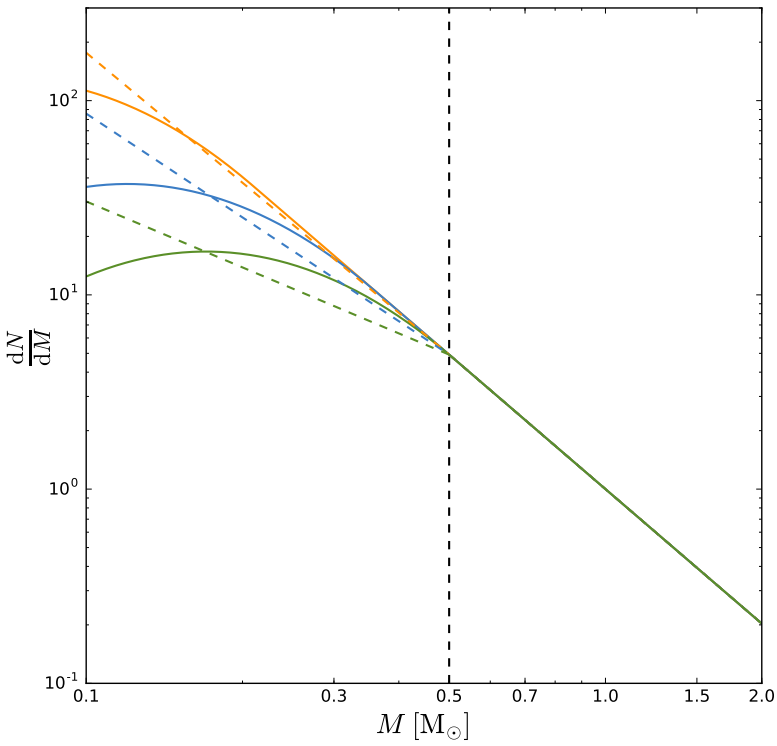

Figure 4.18 – Schematic overview showing that a decreasing low-mass slope of a double power law IMF is consistent with a shift in the characteristic mass of a log-normal IMF below 1 M$_\odot$. These plots are based on a Chabrier IMF (Chabrier 2003) with characteristic masses of 0.17, 0.12, 0.07 M$_\odot$.

as a function of velocity dispersion, indicating that for ETGs with a higher velocity dispersion there is more stellar mass per unit luminosity than for ETGs with lower velocity dispersions due to the steepening of the IMF.

For comparison we also show the dynamical and spectroscopic mass-to-light ratios of the ATLAS^{3D} sample of ETGs (Cappellari et al. 2011) as presented in Cappellari et al. (2013). The dynamical (total) M/L values of the ATLAS^{3D} sample show a much larger spread than the spectroscopic M/L values, which might be related to the fact that the dynamical values are derived independently from stellar evolutionary effects: this might explain the broader range of dynamical M/L values at fixed velocity

dispersion (McDermid et al. 2014). The trends that we find, in particular for `run10`, show good agreement with the lower limits of the dynamically derived M/L-values of the ATLAS^{3D} sample. As the dynamical modelling also includes the mass contribution from dark matter, this is what one would expect.

The spectroscopic M/L values that we find are significantly smaller than the spectroscopic M/L values of the ATLAS^{3D} sample. The spectroscopic M/L values of the ATLAS^{3D} sample are based on a smooth non-parametric SFH and a fixed Salpeter IMF. Since we find IMFs that are flatter than Salpeter for the lower velocity dispersion bins, we also find lower mass-to-light ratios. For the highest velocity dispersion bins, we find IMFs closer to Salpeter, and for these bins our results are consistent with the spectroscopic M/L values of the ATLAS^{3D} sample. The differences between the spectroscopic results by McDermid et al. (2014) and our results may also be explained partly by the different apertures that were used. The SDSS spectra that we analyse only cover the central regions of the galaxies, whereas the ATLAS^{3D} analysis is based on integral-field data out to one effective radius. As a consequence, positive age gradients (i.e. increasing age as a function of radius) in ETGs (e.g., La Barbera et al. 2012) imply that the stellar populations of the SDSS spectra are on average younger than the stellar populations of the ATLAS^{3D} data. Therefore, as a consequence of such an age gradient we will find lower M/L values due to the higher luminosities of the younger populations.

4.6 Summary and discussion

In this paper we have performed a fully hierarchical Bayesian analysis of a set of stacked SDSS spectra of ETGs to infer the shape of their IMF and how it changes as function of their stellar velocity dispersion. The inference is done via full-spectrum fitting with a CSP model that includes a nearly free-form IMF shape, regularized via a single or broken power law, and multiple SSPs with varying ages and metallicities. We also allow the abundance patterns to change by varying elements for which the abundance ratio in ETGs is expected to be different than in the Milky Way. We combine this with an improved library of stellar templates and various evolutionary-track models to construct the CSP models. We also allow the error covariance matrix to vary, and test various priors for robustness of the results. All model parameters are obtained through a nested sampling approach and

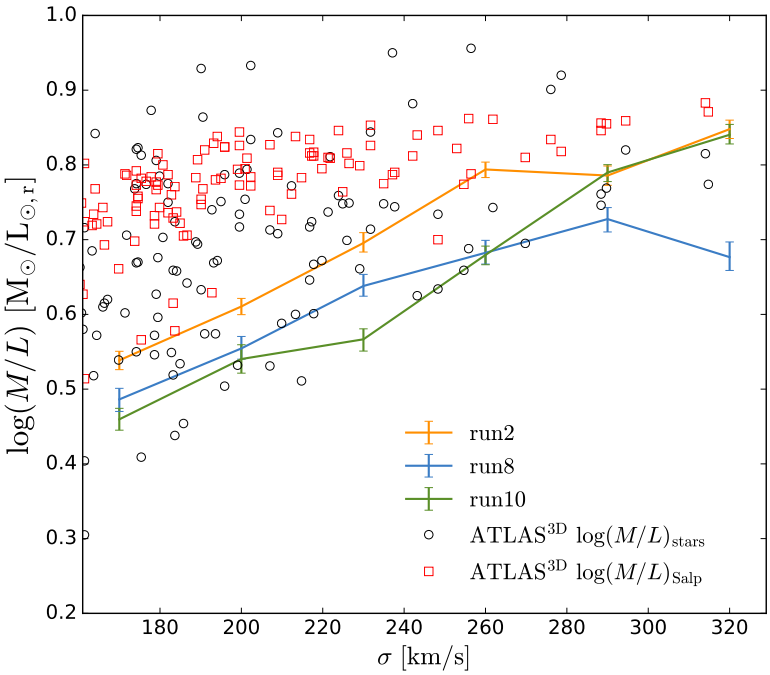

Figure 4.19 – Mass-to-light ratios as a function of velocity dispersion corresponding to the results of `run2`, `run8` and `run10`. Also shown are the dynamical $(M/L)_{\rm stars}$ and spectroscopic $(M/L)_{\rm Salp}$ mass-to-light ratios of the ATLAS$^{\rm 3D}$ sample (Cappellari et al. 2011) derived by Cappellari et al. (2013).

discrete choices (e.g., the set of stellar templates, evolutionary tracks, IMF priors and regularization schemes) are objectively ranked via the Bayesian evidence. This approach ensures that the results that we obtain are robust and avoids, as much as possible, subjective choices.

We extend the model of DTK16 and DTKPS18 and introduce local covariance structures that allow us to add extra covariance to emission lines and telluric regions to better model uncertain spectral regions. Furthermore, we include response functions of Mg, Ca, Si, Ti and Na in the model to account for the presumed variations in the abundance pattern of ETGs with respect to the Milky Way. These response functions are based on SSPs from Conroy & van Dokkum (2012a) and derived on the basis of a Salpeter IMF.

The framework that we have developed can in principle be combined with any set of stellar templates. In this work, the stellar templates that we use are based on the MIX library. The MIX library is a combination of two stellar libraries: MILES and XSL. For every spectrum in the MIX library, the starting point is a VIS spectrum of XSL that we shift to rest frame and correct for extinction. We extend each of these VIS arm spectra to the blue with an interpolated MILES spectrum. To create a consistent interpolator, we run an optimization procedure where we allow for variations in the extinction correction and small variations in the stellar parameters. Finally, we combine the MIX interpolator with two different sets of isochrones to create two sets of stellar templates: the MIX stellar population models.

As a next step, we apply the combination of the hierarchical Bayesian framework and the MIX stellar population models to the stacked SDSS spectra. We consider different runs of the model in which we investigate the effect of various model ingredients, including different combinations of response functions, different isochrones, different regularization schemes and different parameterizations of the IMF prior.

Our main results and conclusions are the following:

- We infer that even high SNR (∼200-1100) stacked SDSS spectra of ETGs can be robustly modelled with at most three SSP components over a wide ($\sigma = 170 - 320$ km/s) mass range. This is in agreement with extensive tests in DTKPS18 on mock spectra with extended SFHs. Some residuals are left between the data and the model, which we model with a flexible covariance matrix, but these are unlikely associated with an un-modelled SSP and more likely due to incompleteness of the set of templates. In particular the sodium feature remains hard to model as found in previous studies.

- We have investigated the effect of various response functions on the relation between IMF slope and velocity dispersion that we recover with our model. For a single power law IMF parameterization and when the response functions of Mg are included, we find an IMF slope that increases from $\alpha \sim 2.1$ for $\sigma = 170\,\mathrm{km\,s^{-1}}$ to $\alpha \sim 2.6$ for $\sigma = 320\,\mathrm{km\,s^{-1}}$. This result is consistent with the IMF slope–velocity dispersion relatation found by Spiniello et al. (2014) and derived with a completely independent code. There is decisive evidence for including the Mg response functions in all the fits and when these response functions are not included the inferred IMF slope–velocity

dispersion relation is biased to a higher value by $\Delta\alpha \sim 0.2$. Including Ca, Si or Ti in combination with Mg does not change the scale of the relation significantly, but including Na decreases the trend by $\Delta\alpha \sim 0.2 - 0.3$. In particular this last result may be affected by how we model the response functions.

- When using a double power law IMF parameterization, we find a high-mass slope that is approximately constant as a function of velocity dispersion and slightly steeper than the Salpeter-value. We find a low-mass slope α_1 that increases as a function of velocity dispersion from approximately the Kroupa-value to the Salpeter-value. The only exception to this are the results that we obtain by including the response functions of Na which suggest flatter high-mass slopes that decrease as a function of velocity dispersion and significantly steeper low-mass slopes. However, with the current setup of the model drawing conclusions from the results that include the Na response functions is not possible. Only when using IMF-dependent response functions instead of SSP based response functions would allow us to disentangle the effect of sodium enhancement from variations in the IMF.

- In most cases the double power law fit is preferred over the single power law fit. For those cases where a single power law IMF fit is preferred, the inferred double power law IMF is actually relatively close to a single power law and therefore the model prefers the most simple model which is the single power law. The results that we find are consistent with a lower characteristic mass of a log-normal IMF for high velocity dispersion ETGs which according to theoretical models could be related to a mode of star formation that is characterized by more turbulence.

Having compared stacked SDSS spectra of ETGs to a wide range of CSP models, varying the IMF shape, ages, metallicities, abundance patterns, and evolutionary tracks, for a varying number of SSPs, we summarize our two major results:

1. There is decisive evidence that the single power law IMF slope increases with galaxy mass over $\sigma_* = 170 - 320\,\mathrm{km\,s^{-1}}$ from $\alpha = 2.1$ to $\alpha = 2.6$, robust against significant changes in the models. This

result is nearly independent from the abundance pattern variations, which typically have impacts of $\Delta\alpha \lesssim 0.2$.

2. There is strong evidence that the IMF slope above $0.5\,M_\odot$ is approximately constant and slightly higher than the Salpeter IMF slope ($\alpha = 2.35$) over the same galaxy mass range, when using a double power law IMF model, but that the IMF slope varies with galaxy mass below $\sim 0.5\,M_\odot$ from being consistent with a Kroupa IMF at the lowest masses ($\alpha \sim 1.3$) to being consistent with a Salpeter IMF for the highest mass ETGs.

Based on the above two major results, we conclude that the IMF shape appears *universal* in shape above $\sim 0.5\,M_\odot$, and *non-universal* below $\sim 0.5\,M_\odot$. This appears to reconcile some of the apparent discrepancies between various results in the literature.

Appendix 4.A SSP-based response functions

In this work we use response functions derived for SSPs, which we interpolate as a function of age and metallicity. These response functions are derived on the basis of an assumed Salpeter IMF. To investigate the consequences of this assumption, we repeat `run2`, `run5` and `run6`, only now with response functions that are derived on the basis of a Kroupa IMF.

Figure 4.20 shows the IMF slope–velocity dispersion relation that we obtain for `run2`, `run5` and `run6` with both the Salpeter and the Kroupa response functions. For Mg, there is a small bias in the reconstructed IMF slopes if we use the Kroupa response functions, but the trend that we recover is very similar. The deviations that we find for Ti appear to be more random than for Mg, even though the values of [Ti/Fe] are relatively small. This suggests that the Ti response functions might affect spectral regions that are also sensitive to the IMF, and therefore we have to be very careful when interpreting these results.

With the exception of `SDSS290`, the inferred IMF slopes with the Na response functions seem to become more biased for higher velocity dispersions. Since higher velocity dispersion galaxies have higher values of [Na/Fe] (see Figure 4.13) and NaD is known to be sensitive to IMF variations, this result is not too surprising. It remains difficult, however, to disentangle the effect of the response functions from variations in the IMF. In principle this should be possible because abundance variations in

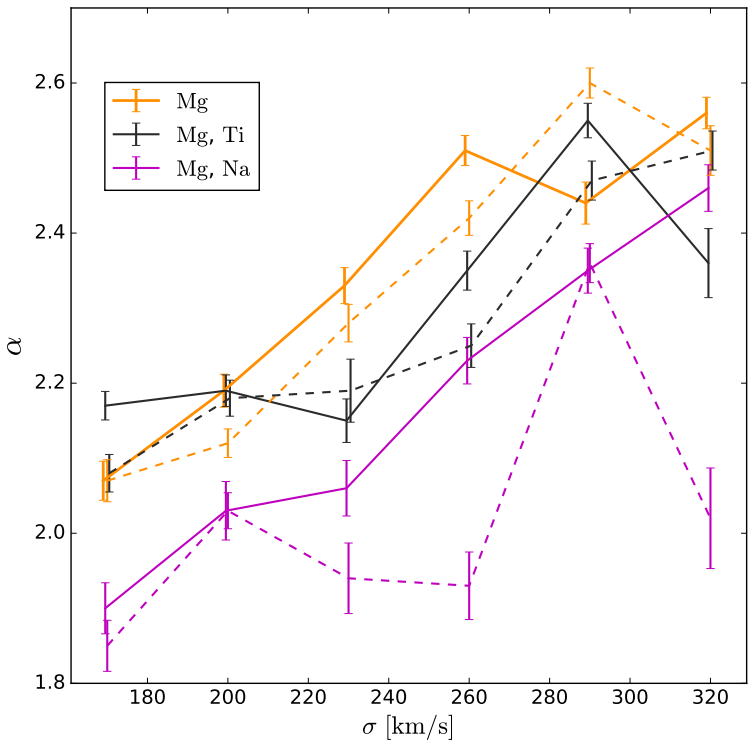

Figure 4.20 – IMF slope–velocity dispersion relation that we infer by using different response functions to account for abundance variations in Mg, Ti and Na. The solid lines are derived by using the SSP-based response functions with a Salpeter IMF and the dashed lines are derived by using the SSP-based response functions with a Kroupa IMF.

Na cause a different spectral response than variations in the IMF does. One of the most important features that allows us to distinguish between IMF variations and abundance variations in Na is the NaI-feature around 8190 Å, but unfortunately this line lies in a strong telluric region which complicates the situation even further. Moreover, we have to keep in mind that the effects of IMF variations on the spectrum are subtle. Therefore, if we want to disentangle this from abundance variations in Na, we have to be sure that we use the correct response functions. In principle this implies that we should apply individual response functions to each stellar template. However, at the moment we only have SSP-based response

functions available to us. The fact that the inferred IMF slopes are in general much closer to a Salpeter IMF than a Kroupa IMF justifies the use of the Salpeter-based response functions throughout this work.

References

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403
Barnabè, M., Spiniello, C., Koopmans, L. V. E., et al. 2013, MNRAS, 436, 253
Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339
Bochanski, J. J., West, A. A., Hawley, S. L., & Covey, K. R. 2007, AJ, 133, 531
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000
Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, MNRAS, 413, 813
Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2013, MNRAS, 432, 1862
Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2012, Nature, 484, 485
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Cenarro, A. J., Gorgas, J., Vazdekis, A., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 339, L12
Chabrier, G. 2003, PASP, 115, 763
Chen, Y., Girardi, L., Bressan, A., et al. 2014, MNRAS, 444, 2525
Chen, Y., Trager, S., Peletier, R., & Lançon, A. 2011, Journal of Physics Conference Series, 328, 012023
Clauwens, B., Schaye, J., & Franx, M. 2016, MNRAS, 462, 2832
Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, ApJ, 780, 33
Conroy, C. & van Dokkum, P. 2012a, ApJ, 747, 69
Conroy, C. & van Dokkum, P. G. 2012b, ApJ, 760, 71
Conroy, C., van Dokkum, P. G., & Villaume, A. 2017, ApJ, 837, 166
Dabringhausen, J., Hilker, M., & Kroupa, P. 2008, MNRAS, 386, 864
Downes, D. & Solomon, P. M. 1998, ApJ, 507, 615
Dries, M., Trager, S. C., & Koopmans, L. V. E. 2016, MNRAS, 463, 886
Dries, M., Trager, S. C., Koopmans, L. V. E., Popping, G., & Somerville, R. S. 2018, MNRAS, 474, 3500
Faber, S. M. & French, H. B. 1980, ApJ, 235, 405
Feroz, F. & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, ArXiv e-

prints
Ferré-Mateu, A., Vazdekis, A., & de la Rosa, I. G. 2013, MNRAS, 431, 440
Ferreras, I., La Barbera, F., de la Rosa, I. G., et al. 2013, MNRAS, 429, L15
Fontanot, F., De Lucia, G., Hirschmann, M., et al. 2017, MNRAS, 464, 3812
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gorgas, J., Faber, S. M., Burstein, D., et al. 1993, ApJS, 86, 153
Graves, G. J. & Faber, S. M. 2010, ApJ, 717, 803
Greve, T. R., Papadopoulos, P. P., Gao, Y., & Radford, S. J. E. 2009, ApJ, 692, 1432
Hauschildt, P. H., Allard, F., & Baron, E. 1999, ApJ, 512, 377
Hennebelle, P. & Chabrier, G. 2008, ApJ, 684, 395
Hopkins, P. F. 2012, MNRAS, 423, 2037
Hopkins, P. F. 2013, MNRAS, 433, 170
Jeffreys, H. 1961, Oxford Univ. Press
Jeong, H., Yi, S. K., Kyeong, J., et al. 2013, ApJS, 208, 7
Johansson, J., Thomas, D., & Maraston, C. 2012, MNRAS, 421, 1908
Kaviraj, S., Khochfar, S., Schawinski, K., et al. 2008, MNRAS, 388, 67
Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, A&A, 501, 1269
Kroupa, P. 2001, MNRAS, 322, 231
Kroupa, P., Tout, C. A., & Gilmore, G. 1993, MNRAS, 262, 545
Krumholz, M. R. 2011, ApJ, 743, 110
Krumholz, M. R. 2014, Phys. Rep., 539, 49
Kurucz, R. L. 1970, SAO Special Report, 309
Kurucz, R. L. 1993, SYNTHE spectrum synthesis programs and line data
La Barbera, F., Ferreras, I., de Carvalho, R. R., et al. 2012, MNRAS, 426, 2300
La Barbera, F., Ferreras, I., Vazdekis, A., et al. 2013, MNRAS, 433, 3017
La Barbera, F., Vazdekis, A., Ferreras, I., et al. 2017, MNRAS, 464, 3597
La Barbera, F., Vazdekis, A., Ferreras, I., et al. 2016, MNRAS, 457, 1468
Lançon, A. & Mouhcine, M. 2002, A&A, 393, 167
Lançon, A. & Wood, P. R. 2000, A&AS, 146, 217
Larson, R. B. 1998, MNRAS, 301, 569
Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. 2004, A&A, 425, 881
Lyubenova, M., Martín-Navarro, I., van de Ven, G., et al. 2016, MNRAS,

463, 3220
MacKay, D. J. C. 1992, Neural Computation, 4, 415
Maraston, C. 2005, MNRAS, 362, 799
Marigo, P., Bressan, A., Nanni, A., Girardi, L., & Pumo, M. L. 2013, MNRAS, 434, 488
Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883
Martín-Navarro, I., Vazdekis, A., La Barbera, F., et al. 2015, ApJL, 806, L31
McDermid, R. M., Alatalo, K., Blitz, L., et al. 2015, MNRAS, 448, 3484
McDermid, R. M., Cappellari, M., Alatalo, K., et al. 2014, ApJL, 792, L37
Monachesi, A., Trager, S. C., Lauer, T. R., et al. 2012, ApJ, 745, 97
Mouhcine, M. & Lançon, A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 192, Spectrophotometric Dating of Stars and Galaxies, ed. I. Hubeny, S. Heap, & R. Cornett, 303
Newman, A. B., Smith, R. J., Conroy, C., Villaume, A., & van Dokkum, P. 2016, ArXiv e-prints
Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2014, Protostars and Planets VI, 53
Padoan, P. & Nordlund, Å. 2002, ApJ, 576, 870
Padoan, P., Nordlund, A., & Jones, B. J. T. 1997, MNRAS, 288, 145
Park, J., Jeong, H., & Yi, S. K. 2015, ApJ, 809, 91
Press, W. H. & Schechter, P. 1974, ApJ, 187, 425
Prugniel, P., Vauglin, I., & Koleva, M. 2011, A&A, 531, A165
Rosenfield, P., Marigo, P., Girardi, L., et al. 2016, ApJ, 822, 73
Sale, S. E., Drew, J. E., Barentsen, G., et al. 2014, MNRAS, 443, 2907
Salpeter, E. E. 1955, ApJ, 121, 161
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Schiavon, R. P. 2007, ApJS, 171, 146
Schiavon, R. P., Barbuy, B., Rossi, S. C. F., & Milone, A. 1997a, ApJ, 479, 902
Schiavon, R. P., Barbuy, B., & Singh, P. D. 1997b, ApJ, 484, 499
Schlafly, E. F. & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Sharma, K., Prugniel, P., & Singh, H. P. 2016, A&A, 585, A64
Smith, R. J., Lucey, J. R., & Conroy, C. 2015, MNRAS, 449, 3441
Spiniello, C., Trager, S., Koopmans, L. V. E., & Conroy, C. 2014, MNRAS, 438, 1483

Spiniello, C., Trager, S. C., & Koopmans, L. V. E. 2015, ApJ, 803, 87
Spiniello, C., Trager, S. C., Koopmans, L. V. E., & Chen, Y. P. 2012, ApJL, 753, L32
Thomas, D., Maraston, C., & Bender, R. 2003, MNRAS, 343, 279
Thomas, D., Maraston, C., Bender, R., & Mendes de Oliveira, C. 2005, ApJ, 621, 673
Tinsley, B. M. 1968, ApJ, 151, 547
Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000a, AJ, 120, 165
Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000b, AJ, 119, 1645
Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, ApJ, 709, 1195
Vazdekis, A., Cenarro, A. J., Gorgas, J., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 340, 1317
Vazdekis, A., Coelho, P., Cassisi, S., et al. 2015, MNRAS, 449, 1177
Vazquez-Semadeni, E. 1994, ApJ, 423, 681
Vernet, J., Dekker, H., D'Odorico, S., et al. 2011, A&A, 536, A105
Westmoquette, M. S., Smith, L. J., Gallagher, III, J. S., et al. 2007, ApJ, 671, 358
Wing, R. F. & Ford, Jr., W. K. 1969, PASP, 81, 527
Worthey, G. 1998, PASP, 110, 888
Worthey, G., Faber, S. M., Gonzalez, J. J., & Burstein, D. 1994, ApJS, 94, 687
Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45

Chapter 5

Conclusions and future outlook

In this thesis, we developed a hierarchical Bayesian framework for stellar population synthesis. We first developed the framework for single stellar populations (SSPs) and demonstrated its viability by applying it to a set of mock SSPs (Chapter 2). Then we extended the model to composite stellar populations (CSPs) and we included a number of additional features in the model such as a multiplicative polynomial, an adaptive covariance matrix and a parameterized version of the code that allows for a much faster sampling procedure. We applied this extended version of the code to a set of mock CSPs with star formation histories based on semi-analytic models and demonstrated the importance of including multiple SSPs in the fit (Chapter 3). As a final step, we extended the model with response functions and local covariance structures and combined it with the MIX-stellar populations models. Using a variety of different model ingredients, we applied the model to a set of stacked *SDSS* spectra. We showed that there exists a correlation between the velocity dispersion of the galaxies and the inferred IMF parameters.

5.1 Results, chapter by chapter

In this section we shortly summarize the results of the scientific chapters.

***Chapter 2**: A hierarchical Bayesian approach for reconstructing the Initial Mass Function of Single Stellar Populations.*

- We developed a hierarchical Bayesian framework with a number of different layers for inferring the IMF of SSPs. At the innermost layer, we use a regulated linear inversion that enables us to determine the individual contributions of stellar templates to the spectrum of the SSP. To regulate this inversion, we use a prior IMF from which the model is allowed to deviate when demanded by the data. At a more outer layer of the model, the free parameters of the IMF prior parameterization are sampled with Markov Chain Monte Carlo sampling techniques. At the outermost level of the model (i.e. the second level of inference), different model ingredients can be compared on the basis of the Bayesian evidence.

- As a first test of the method we apply it to a set of mock SSPs and we show that we can reconstruct the input parameters of these SSPs.

- For old SSPs (∼13 Gyr) it is easier to constrain the low-mass end of the initial mass function (IMF) than for young SSPs (∼3 Gyr) for a fixed SNR of the spectrum. The reason for this is that the younger the SSP, the more light is emitted by stars with $M \geq 0.5\ \mathrm{M}_{\odot}$. Effectively, this decreases the SNR of stars with $M < 0.5\ \mathrm{M}_{\odot}$.

- For bottom-heavy IMFs ($\alpha \approx 3.0$) it is easier to constrain the low-mass end of the IMF than it is for bottom-light IMFs (i.e. a Kroupa IMF). The reason for this is that a bottom-heavy IMF contains more low-mass stars, and therefore the relative contribution of low-mass stars to the integrated spectrum is higher, for a fixed SNR.

- For a double power law parameterization of the IMF prior where the two slopes are not constrained, there is a degeneracy between the low-mass slope and the high-mass slope. An increasing low-mass slope prefers a decreasing high-mass slope and vice versa. A possible explanation for this is the great spectral similarity between low-mass main sequence stars and K and M giants.

- Systematic uncertainties may affect the inference of the IMF. When we apply our model to a set of MILES SSPs, the IMF slopes that we infer are biased. The origin of this bias are systematic uncertainties between the MILES models and our models. On the other hand, when we use stellar templates created with the MILES spectral interpolator, most of the bias is removed. The spectral interpolator is therefore an important source of systematic uncertainty.

- Increasing the signal-to-noise ratio of the spectrum helps to determine better constraints on the IMF, but not if the error budget is dominated by systematic uncertainties.

- To constrain the low-mass end of the IMF for the mock SSPs, the red part of an optical spectrum is required (i.e. $\lambda \approx 5000 - 7400\,\text{Å}$). Given systematic uncertainties in data, the MILES wavelength range ($\lambda \approx 3500 - 7400\,\text{Å}$) alone is currently insufficient to constrain a double power law IMF parameterization.

- Bayesian model comparison is an important feature of the model that we developed. It allows us to objectively compare different ingredients of SPS models on the basis of the evidence.

***Chapter 3**: Hierarchical Bayesian inference of the Initial Mass Function in Composite Stellar Populations.*

- Semi-analytic models are an ideal tool for creating mock spectra of CSPs with realistic star formation histories (SFHs) that can include increasing phases, decreasing phases and bursts of star formation.
- A CSP can be modelled as the superposition of a given number of SSPs. The optimal number of SSPs may be determined on the basis of the Bayesian evidence, and is typically 2 SSPs for SNR=75, 2-3 SSPs for SNR=150 and 5 SSPs for SNR=300.
- When only one SSP is included in the fit to a CSP, the inferred IMF slope may be biased to a value that is higher than the true slope. Most of this bias is already removed when two SSPs are included instead of one.
- Although the inferred IMF slopes may be biased when using only one SSP in the case where multiple SSPs are present, the inferred trend between IMF slope and velocity dispersion is nearly independent of the number of SSPs used to characterize the CSP. Including more SSPs changes the absolute scale of these trends, but less so the slope of the trend.

***Chapter 4**: Hierarchical Bayesian inference of the IMF with the MIX-stellar population models.*

- The MIX-stellar population models allow us to extend the CSP models to ∼10,000 Å and cover more important IMF sensitive features. Combined with the good coverage of the Hertzsprung-Russel diagram by the X-shooter Spectral Library (including the additional M-dwarfs), this makes the MIX-stellar populations models well suited for inference of the IMF in unresolved stellar populations.
- We test the effect of various response functions in the fit of a set of stacked *SDSS* galaxies. There is decisive evidence in favour of including the response function of Mg in the fit. This is consistent with previous studies that show that ETGs are alpha-enhanced. The abundance variations that we find for Ca, Si and Ti are significantly smaller than for Mg and the effect on the evidence is much smaller. Including these response functions seems to make a difference only

for the highest velocity dispersion galaxies, whereas for the lowest velocity dispersion galaxies we find abundance variations close to zero for Ca, Si and Ti. This might be related to a different typical formation history for low and high mass ETGs and the general trend that ETGs with a higher velocity dispersion show stronger variations in the alpha elements (e.g. Trager et al. 2000; Johansson et al. 2012; Conroy et al. 2014).

- The fit to the data that we obtain with our models is in general good, when including an additional global covariance to account for systematic uncertainties and local covariance structures to account for emission lines and strong telluric regions. The most prominent exception to this is the NaD-feature around 5890 Å. To improve the fit in this wavelength region, we test the inclusion of response functions for Na. For all the spectra, we find decisive evidence in favour of including these response functions. However, the trends that we find when using the Na response functions go against all the other trends that we find, which is why we suspect that there is an issue with the results obtained with the Na response functions, and therefore, these results should be considered with caution. This might be related to a degeneracy between the response functions of Na and the inferred IMF parameters. The response functions that we use are calculated for SSPs with a fixed IMF. To resolve this degeneracy requires the use of individual response functions for all the stellar templates, but these response functions are not available to us at the moment.

- All of the results that we obtain by applying our model to a set of stacked *SDSS* spectra of ETGs show that there is a positive correlation between IMF slope and velocity dispersion when using a single power law IMF parameterization and including different response functions. These results agree well with previous results in the literature, but we considered a much wider range of model ingredients which make the results more robust. Nevertheless, it is hard to establish the absolute scale of this trend and this would requires a more accurate modelling of the response functions. In particular the response functions of Na significantly change the scale of the trend.

- Using a different set of isochrones may affect the inferred IMF parameters. We can compare different sets of isochrones on the basis of the Bayesian evidence and select the one that results in the highest evidence.

- A different regularization scheme may result in similar reconstructed IMF prior parameters but different mass-to-light ratios. For a flexible regularization scheme, it is relatively easy for the model to deviate from the prior IMF. Those deviations may result in different mass-to-light ratios than predicted by the prior IMF.

- The evidence for a double power law IMF prior parameterization is in most cases higher than for a single power law IMF prior parameterization. As for the fits with the single power law IMF prior parameterization, the results are significantly different when the response functions of Na are included. Excluding the Na response functions, the results obtained with a double power law parameterization of the IMF prior show a high-mass slope that is approximately constant as a function of velocity dispersion and slightly steeper than the Salpeter-value. The low-mass slope increases as a function of velocity dispersion from approximately the Kroupa-value to the Salpeter-value.

5.2 Main conclusions

The main topic of this thesis has been the construction of a Bayesian model for population synthesis with a specific focus on (the low-mass end of) the IMF. On the basis of two questions defined in the introduction of this thesis, we now present the main conclusions of the work presented in this thesis.

How can one reliably measure the distribution of stellar masses in unresolved galaxies?

A population synthesis model contains many different ingredients. All of these ingredients and the choice that one makes a priori on which ingredients to use or on whether to include a particular ingredient or not, may potentially affect the inference of the IMF.

A reliable determination of the IMF first of all requires high-quality numerical models of stellar evolution to provide the parameters of the stellar

templates in the form of isochrones. The stellar library that is used is preferably empirical, provides a good coverage of the HR diagram, has a good wavelength coverage, is well-calibrated and the stellar parameters of the stars in the library have been determined accurately. Moreover, a reliable interpolation mechanism is required to convert the spectra in the library to a set of stellar templates on the basis of the stellar parameters of the isochrones. The choice of IMF parameterization can also limit the stellar mass distributions available to the model. Therefore, it is preferable to use more flexible models that depend less on the parameterization of the IMF. At least one should consider and compare different parameterizations of the IMF.

Apart from these basic ingredients, full spectrum fitting requires a multiplicative polynomial to model smooth differences between the continua of the data and the model spectrum as a consequence of extinction, issues with the telluric correction and issues with the flux calibration. Broadening of the spectra due to the velocity dispersion of the stars also has to be included in the model. Our results show that ETGs should not be modelled as SSPs and that abundance variations of these galaxies with respect to the stars in the Milky Way (and hence the stars in the stellar library) should properly be taken into account.

As can be seen from these results, inference of the IMF in unresolved stellar populations is a complicated problem that requires accurate models (to the level of a percent or less in an optical spectrum to detect the small spectral changes resulting from IMF variations). Testing different model ingredients and comparing these ingredients on the basis of the Bayesian evidence provides a powerful way to select the appropriate ingredients. Model comparison is, however, open-ended and the ingredients that we test in this thesis are not exhaustive. Therefore, future research might lead to the conclusion that additional or improved ingredients are required.

Is the stellar mass distribution the same in all galaxies?

In Chapter 4 we applied the model that we developed in this thesis to a set of stacked *SDSS* spectra of ETGs binned by velocity dispersion. We tested a variety of different model ingredients that we assessed on the basis of the Bayesian evidence. We find an IMF that is not universal, in which the inferred IMF parameters change as a function of velocity dispersion. The results that we obtain show that ETGs with a higher mass/velocity dispersion have an IMF that is more bottom-heavy than

ETGs with a lower mass/velocity dispersion, confirming earlier results in the literature (e.g. Spiniello et al. 2012; Conroy & van Dokkum 2012; La Barbera et al. 2013; Martín-Navarro et al. 2015). With respect to the shape of the IMF, we find strong preference for a double power law IMF parameterization as compared to a single power law IMF parameterization. For the double power law IMF parameterization, we find that the high-mass slope is approximately constant and that the low-mass slope varies as a function of velocity dispersion, with more bottom-heavy IMFs for higher velocity dispersions. There are various sources of uncertainty in the model, such as the response functions that we use to account for variable abundance patterns, that complicate the inference of the absolute scale of the IMF variations and the exact shape of the IMF. In the future outlook below we include some suggestions on how to improve the constraints on the IMF presented in this work.

Although the results presented in this thesis show the existence of an empirical relation between the IMF and the velocity dispersion of ETGs, the global velocity dispersion of a galaxy is unlikely directly responsible for driving variations of the IMF. The IMF is the outcome of the physics that describes the complicated process of star formation, and variations of the IMF are expected to be related to a change in the local properties of the star forming interstellar medium. One of these properties is the Mach number $\mathcal{M}$ that characterizes the turbulence of the interstellar medium (e.g. Hennebelle & Chabrier 2008). In Chapter 4 we discussed how a change in $\mathcal{M}$, as a consequence of different typical circumstances of the interstellar medium, may result in a shift of the characteristic mass of a log-normal IMF (see also Chabrier 2003; Offner et al. 2014), providing a possible physical explanation for the non-universality of the IMF. We showed that the results that we obtain in Chapter 4, when using a double power law IMF parameterization, are consistent with this theoretical model.

5.3 Future outlook

The model that we developed in this thesis provides a solid statistical framework for population synthesis studies in general and inference of the IMF in unresolved stellar populations in particular. I conclude this thesis with some ideas for future research and improvements of the model that will help to further constrain the non-universal IMF.

- **Reduce number of templates in SSP**

  The number of stars in an SSP is determined by the isochrone that is used. As a consequence the number of stellar templates in an SSP can be quite high. This in turn significantly affects the computational time scales for evaluating the model. In Chapter 3 we described a binning procedure to reduce the number of stellar templates. We suspect that the number of templates in an SSP may be reduced even further but this requires a more systematic and careful analysis of the effect of the binning procedure on the inferred (IMF) properties. The advantage of reducing the number of templates used to represent an SSP is that it will significantly speed up the model, which would then allow us to include for example more combinations of response functions in the model.

- **Improve response functions**

  The response functions that we used in this thesis are based on SSPs under the assumption of a fixed IMF. However, the response functions and the IMF are not independent. Therefore, if we want to disentangle variations of the IMF from abundance variations, it is important to apply star-based response functions to the individual stellar templates. Once these response functions are available, it will be rather time-consuming to take this into account in the model so in that respect we relate this to the previous point by stating that it is important to find ways to increase the speed of the model.

- **Use full wavelength range of XSL**

  In the application to the data presented in this thesis, we use the MIX-stellar populations models. These models only use the VIS arm of XSL. The main reasons for not using the UVB arm of XSL are the uncertain extinction correction and the uncertain shape of the UVB/VIS spectra in the dichroic-region. Since we applied the model to *SDSS* spectra, the NIR arm was not required in this work. Nevertheless, the NIR arm contains additional IMF sensitive features and using the full UVB-VIS-NIR wavelength range of XSL provides a stellar library that is more consistent. Therefore, once XSL is completely finished and correctly calibrated, we strongly suggest combining the model that we developed with the full library and applying this combination to a set of ETG spectra observed with X-

shooter to employ the full potential of the wavelength range offered by XSL.

- **Test additional model ingredients**

  The collection of model ingredients that we test in this thesis is a model choice and definitely not exclusive. There are other model ingredient that are worth investigating and for which the model that we developed provides a useful framework. Examples of these ingredients are the value of the low-mass cut-off of the IMF (Barnabè et al. 2013) and allowing different SSPs in a fit to a CSP to have different IMF shapes.

- **Improve modelling of AGB stars**

  The asymptotic giant branch is one of the most uncertain phases of stellar evolution. Yet, these stars may contribute significantly to the light emitted by a stellar population. To convert the stars in a stellar library to a spectrum, the spectra in the library are often interpolated on the basis of three parameters: the effective temperature, the surface gravity and the metallicity. For AGB stars, interpolating a stellar spectrum on the basis of these three parameters is clearly insufficient. Therefore, it is important to further investigate which parameters are required to model the spectra of these stars and what would be the best way to include these stars in population synthesis models.

- **Use Bayesian framework for SFH inference**

  As was suggested in Chapter 3, the framework that we developed in this thesis may also be used to infer the SFH of galaxies. In that case one would need to assume a fixed IMF and create a set of SSP template spectra under the assumption of this IMF. Instead of stellar templates, the columns of matrix $\mathbf{S}$ are now formed by a set of SSP templates, and instead of a prior on the IMF, one could use a prior on the SFH to regulate a direct inference of the SFH of a galaxy.

References

Barnabè, M., Spiniello, C., Koopmans, L. V. E., et al. 2013, MNRAS, 436, 253
Chabrier, G. 2003, PASP, 115, 763
Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, ApJ, 780, 33
Conroy, C. & van Dokkum, P. G. 2012, ApJ, 760, 71
Hennebelle, P. & Chabrier, G. 2008, ApJ, 684, 395
Johansson, J., Thomas, D., & Maraston, C. 2012, MNRAS, 421, 1908
La Barbera, F., Ferreras, I., Vazdekis, A., et al. 2013, MNRAS, 433, 3017
Martín-Navarro, I., Vazdekis, A., La Barbera, F., et al. 2015, ApJL, 806, L31
Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2014, Protostars and Planets VI, 53
Spiniello, C., Trager, S. C., Koopmans, L. V. E., & Chen, Y. P. 2012, ApJL, 753, L32
Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 120, 165

Appendix A

HBSPS manual

`HBSPS` (Hierarchical Bayesian Stellar Population Synthesis) is a stellar population synthesis code designed for inferring the initial mass function (IMF) and other properties from the spectra of unresolved stellar populations. The code is publicly available at `https://github.com/mdries/HBSPS`. The manual that is included in this thesis refers to version 1.0 of HBSPS. This manual might change with future releases of the code: the latest version of it will be available in the GitHub repository.

The model and several updates to the model are described in Dries et al. (2016), Dries et al. (2018) and Dries et al. (in prep.) (Chapter 2, 3 and 4 of this thesis). Within the current version of the model the following features are included:

- a composite stellar population consisting of multiple single stellar populations (SSP) fits;
- flexible IMFs with single and double power law IMF parameterizations of the IMF prior;
- a multiplicative polynomial for absorbing the continuum mismatch between the data and the model;
- a Gaussian kernel for modelling the velocity dispersion of the stellar population;
- an additional global covariance terms to take into account systematic uncertainties;

- local covariance structures for increasing the error in uncertain spectral regions.

Note that although SSP-based response functions for Mg, Ca, Si, Ti and Na are implemented in the code, the response functions themselves are not included in version 1.0 of `HBSPS`. In practice this implies that the abundance pattern is fixed to approximately solar.

Running `HBSPS` requires a set of stellar templates. You can either use your own set of stellar templates, or you can use the stellar templates that are distributed along with `HBSPS`. The stellar templates that come with `HBSPS` are based on the MILES library (Sánchez-Blázquez et al. 2006) and are calculated for two sets of isochrones: the Parsec isochrones (Bressan et al. 2012) and a set of stitched isochrones based on Conroy & van Dokkum (2012); Baraffe et al. (1998); Dotter et al. (2008); Marigo et al. (2008) and described in Dries et al. (in prep.). For the interpolation of these stellar templates, we use the spectral interpolator described in Dries et al. (2016) which is based on Vazdekis et al. (2003). If you use these stellar templates, please reference the aforementioned papers on which these templates are based. Before you can use these stellar templates, you need to install them as described in Section A.1.

There are two versions of the code: the parameterized version and the full version. In the parameterized version of the code, the model spectrum is compared directly to the data on the basis of a χ^2-value, and the IMF is enforced to be either a single or a double power law IMF. In the full version of the code, the IMF is inferred through a linear inversion that allows deviations from the IMF prior if demanded by the data. We use the parameterized version of the code to infer the velocity dispersion and the ages and metallicities of the SSPs that we fit to the data. We then run the full version of the code with fixed SSPs and velocity dispersion to infer the IMF. The reason for using this sampling procedure is that it is much faster than running the full version of the code with free SSP ages and metallicities but it implicitly assumes that the IMF is only a weak function of age and metallicity. Hence each of the SSPs is assumed to have an identical IMF shape. For more details of the model we refer the reader to the aforementioned papers. In the remainder of this document we describe how to install and use `HBSPS`.

A.1 Installation

HBSPS is written in Python and is developed as a module of the cosmological parameter estimation code CosmoSIS (Zuntz et al. 2015). The advantage of combining HBSPS with CosmoSIS is that it allows to combine HBSPS with a wide variety of sampling techniques and that it provides a clear interface to the sampler. We mostly use Multinest (Feroz & Hobson 2008; Feroz et al. 2009, 2013), since we often want to determine the Bayesian evidence, but changing to a different sampler is straightforward with CosmoSIS.

Before using HBSPS, make sure to install CosmoSIS and its dependencies. We do not describe the installation of CosmoSIS here, which is already extensively described on the CosmoSIS wiki: https://bitbucket.org/joezuntz/cosmosis/wiki/Home. We recommend reading the CosmoSIS wiki and trying some of the examples provided with the code before using HBSPS to get a basic understanding of how to use CosmoSIS. Within HBSPS we use the HDF5 binary data format for large data files (i.e. the stellar templates and the response functions). The Python interface to this data format is provided by the package h5py (http://www.h5py.org), which is required before using HBSPS.

After installing CosmoSIS, h5py and cloning the GitHub directory to a directory of your choice, you are ready to start using HBSPS with your own set of stellar templates. Running HBSPS with the MILES-based stellar templates requires an additional step that is described below.

A.1.1 Installation stellar templates

To install the stellar templates that are distributed with HBSPS, go to the directory MILEStemplates in the HBSPS directory, then run the python script installMILEStemplates.py. The process of interpolating all the stellar templates can be quite time consuming, but it is possible to use multiple processors for this by using the option -n. By default, the stellar templates are not binned. To bin the stellar templates under the assumption of a Salpeter IMF, and by the procedure described in Dries et al. (2018), use the option -b. This binning procedure reduces the computational time scales for running HBSPS significantly. As an example, to install the templates, bin them and use eight processors, one would use the following command

```
python installMILEStemplates.py -b -n 8
```

A.2 Input file

The file containing the input spectrum should be formatted according to the input file expected by `HBSPS`. Currently, `HBSPS` only supports plain text files where the data is ordered into three columns. The first column should represent the wavelength grid of the data, the second column the input spectrum and the third column the error spectrum in the same units as the input spectrum. The wavelength grid of the data and the wavelength grid of the stellar templates are expected to be the same so this might require to rebin the stellar templates to the wavelength grid of the data or vice versa (the file `specBasics.py` contains a function for this). If necessary, `HBSPS` can be changed to deal with other file formats as well.

A.3 Stellar templates

`HBSPS` can only be used in combination with a set of stellar templates. These stellar templates may be created by combining a set of isochrones with a stellar library and a spectral interpolator (see e.g. Dries et al. 2016). Within `HBSPS` the stellar templates are assumed to be in a separate directory where there is one HDF5-file for each SSP in the log age-metallicity grid.

Currently we adopt a log age-metallicity grid with

$$\log t = 9.175 \longrightarrow 10.125 \,\text{and}\, \Delta \log t = 0.025, \tag{A.1}$$

i.e. templates between 1.5 and 13.3 Gyr, and with

$$[\text{Fe/H}] = -1.4 \longrightarrow 0.4 \,\text{and}\, \Delta[\text{Fe/H}] = 0.05. \tag{A.2}$$

This log age-metallicity grid is hardcoded in `HBSPS` so you need to change the code if you want to use a different age-metallicity grid. The HDF5-files corresponding to the individual SSPs in the age-metallicity grid should at least contain the following datasets:

- `mass`: masses (in $M_\odot$) of the stellar templates, make sure that this array is strictly increasing and there should not be multiple templates that have the same mass.
- `massLow`: lower-mass boundaries (in $M_\odot$) of the IMF mass bins for converting IMF to weights and vice versa.

- `massUp`: upper-mass boundaries (in $M_\odot$) of the IMF mass bins for converting IMF to weights and vice versa.
- `spectra`: spectra of the stellar templates in a two-dimensional array where each of the columns represents a stellar spectrum (units should be consistent with the input spectrum).
- `luminosity`: bolometric luminosity of the stellar templates for calculating mass-to-light ratios.

Make sure to use the same names for the datasets as above or update the file `SPSbasics.py` to use your own naming convention. The naming convention for the HDF5-files is as follows:

`templates-logT` + $\log t$ + `-FeH` + [Fe/H] + `.hdf5`

For example, the templates with $\log t = 9.775$ and [Fe/H] $= 0.1$ are stored in the file `templates-logT9.775-FeH0.1.hdf5`. To use your own naming convention, simply update the file `SPSbasics.py`.

A.4 Setup files

To run `CosmoSIS` in combination with the `HBSPS`-modules `HBSPSPar.py` (parameterized version of the code) and `HBSPS.py` (full version of the code) we need to create a `CosmoSIS` parameter file and a `CosmoSIS` values file. We discuss the parameters in these files on the basis of some example files. We first discuss the parameter file of the parameterized version of the code, then the parameter file for the full version of the code and finally we discuss the value files.

A.4.1 Parameterized version of `HBSPS`

In Listing A.1 we include the file `examplePar.ini`, which gives an example parameter file for the parameterized version of the code. The `Python`-module that is required for running this code is `HBSPSPar.py`.

```
[runtime]
sampler = multinest

[multinest]
max_iterations = 50000
live_points = 600
```

```
feedback = True
tolerance = 1.0
update_interval = 200
log_zero = -1e14
multinest_outfile_root = /Users/matthijs/PhD/public_code/HBSPS/
    ↪ output/

[output]
filename = /Users/matthijs/PhD/public_code/HBSPS/output/SSPParsec
format = text

[pipeline]
modules = HBSPSPar
values = /Users/matthijs/PhD/public_code/HBSPS/valuesPar.ini
likelihoods = HBSPSPar
quiet = T
timing = F
debug = F

[HBSPSPar]
file = /Users/matthijs/PhD/public_code/HBSPS/HBSPSPar.py
inputSpectrum = testSSPParsec.txt
templatesDir = MILEStemplates/Parsec/
resFunHDF5 = None
nSSPs = 2
nSlopes = 1
logbcov = -2.0
polOrder = 10
sampleMg = False
sampleCa = False
sampleSi = False
sampleTi = False
sampleNa = False
```

Listing A.1 – Example of parameter file for parameterized version code.

The parameter file consists of different sections that may be summarized as follows

- The section [**runtime**] defines the basic parameters for running the code and should at least specify which sampler should be used.
- For the sampler you choose to use in the [**runtime**] section, there should be a section with the name of the specified sampler: in this

Table A.1 – Free parameters in `CosmoSIS` parameter file for running parameterized version of the code.

parameter	description
`file`	Path to the `Python` file `HBSPSPar.py` that will be used for likelihood evaluation.
`inputSpectrum`	Path to the input spectrum. Note that the input spectrum should be in the format described in Section A.2.
`templatesDir`	Path to the directory with the stellar templates. See Section A.3 for more information on the format of the HDF5-files with the stellar templates.
`resFunHDF5`	Path to the HDF5-file containing the response functions. If sampleMg, sampleCa, sampleSi, sampleTi and sampleNa are all set to `False` then this parameter does not need to be specified.
`nSSPs`	Number of SSPs included in the fit.
`nSlopes`	Number of IMF slopes in the parameterization of the IMF: 1 = single power law IMF, 2 = double power law IMF. Change code for a fixed IMF or variable break mass.
`logbcov`	Logarithm of the additional global covariance that will be used to inflate the covariance matrix to deal with systematic uncertainties. Note that `logbcov` is parameterized such that the new (diagonal) covariance matrix becomes $\mathbf{C}_{\mathrm{D,new}} = \mathbf{C}_{\mathrm{D,old}} \left(1 + 10^{\mathrm{logbcov}}\right)$.
`polOrder`	Order of the multiplicative polynomial. Use `polOrder`=0 for a constant rescaling.
`sampleMg`, `sampleCa` `sampleSi`, `sampleTi` `sampleNa`	Booleans to indicate if abundance variations in Mg, Ca, Si, Ti, Na are modelled or not. Currently not yet fully implemented, so set to `False`.

case we use `Multinest`, so we include the section [**multinest**]. The options that you can specify in this section depend on the sampler. See the wiki of `CosmoSIS` for more information on the different samplers and their options.

- The [**output**] section specifies where to store the output results after the sampling procedure has finished. See the `CosmoSIS` wiki for more details.
- In the [**pipeline**] section the external module(s) that `CosmoSIS` should use for the likelihood evaluation are described. For every module specified here there should be a separate section in the parameter file. In this case we use the external module HBSPSPar for likelihood evaluations, so in the parameter file there should also be a section [**HBSPSPar**]. In this section also the reference to the values file is specified. For more details see the `CosmoSIS` wiki.
- The [**HBSPSPar**] section specifies the options for running the parameterized version of the code. The different options that are available for running the parameterized version of the code are specified in Table A.1.

A.4.2 Full version of `HBSPS`

In Listing A.2 we include an example of a parameter file for the full version of the code. This parameter file is very similar to the parameter file of the parameterized version of the code. The module that is used for likelihood evaluations is HBSPS and therefore the parameter file now has a section [**HBSPS**] instead of the section [**HBSPSPar**]. Within this section the options for running the full version of the code are specified. The different options that are available for running the full version of the code are specified in Table A.2.

```
[runtime]
sampler = multinest

[multinest]
max_iterations = 50000
live_points = 300
feedback = True
update_interval = 200
```

```
log_zero = -1e14
multinest_outfile_root = /Users/matthijs/PhD/public_code/HBSPS/
    ↪ output/

[output]
filename = /Users/matthijs/PhD/public_code/HBSPS/output/SSPParsec
format = text

[pipeline]
modules = HBSPS
values = /Users/matthijs/PhD/public_code/HBSPS/values.ini
likelihoods = HBSPS
quiet = T
timing = F
debug = F

[HBSPS]
file = /Users/matthijs/PhD/public_code/HBSPS/HBSPS.py
inputSpectrum = testSSPParsec.txt
templatesDir = MILEStemplates/Parsec/
resFunHDF5 = None
nSSPs = 2
nSlopes = 1
sigma = 150.0
ageIndices = 38 30
FeHIndices = 28 30
polOrder = 10
sampleMg = False
sampleCa = False
sampleSi = False
sampleTi = False
sampleNa = False
```

Listing A.2 – Example of parameter file for full version of the code.

A.4.3 Values file

In addition to a parameter file, both versions of the code require a "values" file. The values file specifies the free parameters of the model (i.e. those parameters that should be sampled by the sampler) and the (uniform) priors for these parameters. For every free parameter in the model the values file provides three numbers: the first number is the lower boundary

Table A.2 – Free parameters in `CosmoSIS` parameter file for running full version of the code.

parameter	description
`file`	Path to the `Python` file `HBSPS.py` that will be used for likelihood evaluation.
`inputSpectrum`	Path to the input spectrum. Note that the input spectrum should be in the format described in Section A.2.
`templatesDir`	Path to the directory with the stellar templates. See Section A.3 for more information on the format of the HDF5-files with the stellar templates.
`resFunHDF5`	Path to the HDF5-file containing the response functions. If sampleMg, sampleCa, sampleSi, sampleTi and sampleNa are all set to `False` then this parameter does not need to be specified.
`nSSPs`	Number of SSPs included in the fit.
`nSlopes`	Number of IMF slopes in the parameterization of the IMF: 1 = single power law IMF, 2 = double power law IMF.
`sigma`	Velocity dispersion of the stellar population (from parameterized version of the code).
`ageIndices`	Age-indices of SSPs in age-metallicity grid to use in the fit (from parameterized version code).
`FeHIndices`	FeH-indices of SSPs in age-metallicity grid to use in the fit (from parameterized version code).
`polOrder`	Order of the multiplicative polynomial.
`sampleMg`, `sampleCa` `sampleSi, sampleTi` `sampleNa`	Booleans to indicate if abundance variations in Mg, Ca, Si, Ti, Na are modelled or not. Currently not possible, so set to `False`.

for the uniform prior of the parameter, the second number is a starting value for the parameter, and the third number is the upper boundary for the uniform prior of the parameter. Obviously the starting value should lie somewhere between the lower and upper boundary of the prior. In Listing A.3 we include an example of a values file for the parameterized version of the code with two SSPs included in the fit. The values file for the full version of the code has the same structure (but different parameters). Within `CosmoSIS` it is also possible to use non-uniform priors for the model parameters. If you want to use these, the `CosmoSIS` wiki explains how to include them.

A.4.3.1 IMF normalization: parameterized code version

In the parameterized version of the model, the normalization of the IMF that determines the contribution of the different SSPs to the total integrated flux is parameterized in the form of the parameter `lumFrX`. This parameter represent the relative contribution of SSP `X` to total integrated luminosity of the input spectrum. We assume that

$$\sum_{i=1}^{n_{\mathrm{SSPs}}} \mathrm{lumFr}_i = 1.0, \tag{A.3}$$

which implies that the value of `lumFr` for the last SSP can be derived from this equation. Therefore, there is always one `lumFr`-parameter less than the number of SSPs in the values file of the parameterized version of the code which increases the sampling speed significantly. For example, in Listing A.3 two SSPs are included in the fit, but only the parameter `lumFr1` is sampled and `lumFr2` is derived from Equation A.3. In principle, for more than two SSPs, the sum of the relative contributions might become more than one but we include a prior to prevent this (otherwise there would be a degeneracy with the multiplicative polynomial).

```
[parameters]
alpha = 0.0 2.3 4.0
sigma = 140 260 350
FeH1 = -1.4 0.0 0.4
age1 = 1.5 8.0 13.5
lumFr1 = 0.001 0.9 1.0
FeH2 = -1.4 0.0 0.4
age2 = 1.5 8.0 13.5
```

Listing A.3 – Example of values file for parameterized version of the code.

A.5 Running the code

Once you have a parameter file and a values file for either the parameterized version or the full version of the code, running it with `CosmoSIS` is straightforward. If the parameter file is `example.ini`, use

```
cosmosis example.ini
```

to run the code. If the sampler that you use allows for parallel sampling, you can run the sampler in parallel with e.g. 8 cores with

```
mpirun -np 8 cosmosis --mpi example.ini
```

Note that this requires a working MPI environment and the `Python` package `mpi4py`.

A.6 Output files

The minimum output of a sampling procedure is the file specified under `filename` in the section [**output**] of the parameter file. To generate some summary statistics and plots you can use the `postprocess` command that comes with the installation of `CosmoSIS`. If the output file is `output.txt`, simply use

```
postprocess output.txt
```

to generate the summary statistics and plots. Generating the plots may be time-consuming, if you only want to see the statistics use the option `--noplots`.

Note that if you use `Multinest`, the output file is not a posterior sample but a set of weighted samples. If the parameter `multinest_outfile_root` is set in the [**multinest**] section of the parameter file, this will generate the `Multinest` output files. One of these files is `post_equal_weights.dat`, which contains a posterior sample of the sampled parameters.

A.7 Running the pipeline

In principle, the parameterized and the full version of the code may be used independently[1]. However, the idea is to use the parameterized and the full version of the code together. First the parameterized version of the code is used to determine the velocity dispersion and the ages and metallicities of the SSPs. Then the full version of the code is used to sample the IMF-related parameters and the global covariance parameter `logbcov`, with the values of the ages, metallicities and velocity dispersion kept fixed.

To automate this process we have developed a pipeline for running `HBSPS`. The only thing that you need to do before running this pipeline is to set the configuration options in the file `HBSPS.ini`. Then run the pipeline with the command

[1]The full version of the code only supports fixed ages and metallicities of the SSPs and a fixed velocity dispersion but it is relatively easy to change this in the code.

`python HBSPSpipeline.py` *`inputSpectrum`*

where *`inputSpectrum`* is the name of the file with the input spectrum. The pipeline will then first run the parameterized version of the code, determine the best-fitting velocity dispersion and SSP ages and metallicities from this run, and then run the full version of the code with these parameters. Parameter files and values files are created automatically by the pipeline.

In Listing A.4 we include an example of the configuration file `HBSPS.ini`. The different options in this file are summarized in Table A.3.

```
[templates]
templatesDir = /data/users/dries/specInt/XSL/stitched-data/
    ↪ hdf5FilesSDSS/
nSSPs = 3

[IMFprior]
nSlopes = 1
regScheme = 1

[responseFunctions]
sampleMg = True
sampleCa = False
sampleSi = False
sampleTi = False
sampleNa = False
hdf5File = /net/thales/data/users/dries/research/model-v2/data/
    ↪ intResFunctions.hdf5

[polynomial]
order = 10

[covariance]
logbcov = 1.587

[sampling]
outputDir = /net/thales/data/users/dries/research/model-v2/
    ↪ public_code/HBSPS/output/
nCores = 64
livepoints = 300

[priors]
alpha = 0.0 2.3 4.0
alpha1 = 0.0 2.3 4.0
```

```
alpha2 = 0.0 2.3 4.0
norm = -4.0 0.0 2.0
sigma = 140 260 350
age = 1.5 8.0 13.5
FeH = -1.4 0.0 0.4
logbcov = -2.0 0.0 4.0
dex = -0.3 0.0 0.4
```

Listing A.4 – Example of the configuration file `HBSPS.ini` required for running the `HBSPS` pipeline.

A.8 Running a test spectrum

To test whether the code is working correctly, we included two SSP test spectra. The spectrum `testSSPParsec.txt` was created with the Parsec isochrones and the spectrum `testSSPstitched.txt` with the stitched isochrones. Both spectra correspond to an SSP with $t = 13.0\,\mathrm{Gyr}$, [Fe/H] = 0.0 and a Salpeter IMF slope ($\alpha = 2.35$). The signal-to-noise ratio (SNR) of the spectra is SNR = 150. Once the file `HBSPS.ini` has been configured with the correct options (see Section A.7), the test spectra can be run with

```
python HBSPSpipeline.py testSSPParsec.txt/testSSPstitched.txt
```

If you use the original stellar templates with one SSP, you should recover the input parameters of the SSPs.

Table A.3 – Free parameters in `HBSPS.ini`, the configuration file for running the `HBSPS` pipeline.

parameter	description
	[templates]
`templatesDir`	Path to directory with stellar templates.
`nSSPs`	Number of SSPs to use in the fit.
	[IMF prior]
`nSlopes`	Number of slopes in IMF (prior) parameterization: 1 = single power law IMF and 2 = double power law IMF.
`regScheme`	Which regularization scheme should be used in the full version of the code: 1 = identity matrix and 2 = $1/w_0^2$.
	[Response functions]
`sampleMg`, `sampleCa`	Booleans to indicate if response functions of these elements should
`sampleSi`, `sampleTi` `sampleNa`	be sampled.
`hdf5File`	Path to the HDF5-file with the response functions. If no response functions are used this parameter is not necessary.
	[polynomial]
`order`	Order of the multiplicative polynomial.
	[covariance]
`logbcov`	Logarithm of the additional covariance used in the parameterized version of the code.
	[multinest]
`outputDir`	Directory where the output files will be saved.
`nCores`	Number of CPU's to use in the sampling procedure.
`livepoints`	Number of livepoints used by `Multinest` in the sampling procedure.
	[priors]
`alpha...dex`	Priors and starting values for sampled parameters that are used for creating the values files (see Section A.4.3).

References

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127

Conroy, C. & van Dokkum, P. 2012, ApJ, 747, 69

Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89

Dries, M., Trager, S. C., & Koopmans, L. V. E. 2016, MNRAS, 463, 886

Dries, M., Trager, S. C., Koopmans, L. V. E., Popping, G., & Somerville, R. S. 2018, MNRAS, 474, 3500

Feroz, F. & Hobson, M. P. 2008, MNRAS, 384, 449

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, ArXiv e-prints

Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883

Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703

Vazdekis, A., Cenarro, A. J., Gorgas, J., Cardiel, N., & Peletier, R. F. 2003, MNRAS, 340, 1317

Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45

Nederlandse Samenvatting

Sterrenstelsels en de initiële massafunctie

Daar waar het 's nachts nog echt donker wordt, kan men als het helder is een lichtende band aan de hemel ontwaren. Al sinds de oudheid staat deze band bekend als de Melkweg. Toen Galileo Galilei in 1610 voor het eerst de Melkweg bestudeerde met een telescoop, ontdekte hij dat de Melkweg bestaat uit een grote hoeveelheid zwakke sterren die niet zichtbaar zijn met het blote oog.

In de achttiende eeuw publiceerde de Franse kometenjager Charles Messier een lijst van objecten die vaak werden verward met kometen, maar in werkelijkheid geen kometen konden zijn omdat hun positie aan de hemel niet veranderde. Eén van deze objecten was de Andromedanevel. Met behulp van variabele sterren toonde de Amerikaanse astronoom Edwin Hubble in 1923 aan dat de Andromedanevel zich buiten de Melkweg bevindt en daarmee een op zichzelf staand sterrenstelsel is. Inmiddels weten we dat de Melkweg slechts één van miljarden sterrenstelsels is.

Om al deze verschillende sterrenstelsels te rangschikken ontwikkelde Edwin Hubble een classificatieschema voor sterrenstelsels in de vorm van een stemvorkdiagram zoals te zien in Figuur A.1. In dit classificatieschema wordt met name onderscheid gemaakt tussen elliptische sterrenstelsels en spiraalvormige sterrenstelsels. Spiraalvormige sterrenstelsels zijn vaak blauw van kleur, daar waar elliptische sterrenstelsels gekarakteriseerd worden door een geelrode kleur. De typische kleuren van deze twee groepen van sterrenstelsels zijn gerelateerd aan de sterren die we in deze

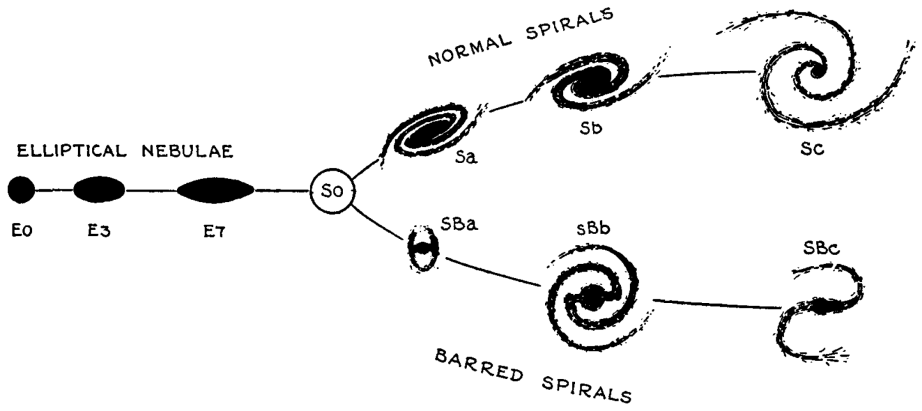

Figure A.1 – Classificatie van sterrenstelsels met behulp van het Hubble stemvorkdiagram. Bij deze classificatie wordt met name onderscheid gemaakt tussen elliptische stelsels en spiraalstelsels. Bron: Hubble (1936).

sterrenstelsels aantreffen. In spiraalvormige sterrenstelsels bevindt zich over het algemeen veel gas, wat er voor zorgt dat in deze sterrenstelsels nog actief sterren worden gevormd. De zwaarste sterren die hierbij gevormd worden domineren het licht van deze sterrenstelsels en geven spiraalstelsels hun blauwachtige kleur. In elliptische stelsels daarentegen, is de hoeveelheid nieuwe sterren die wordt gevormd minimaal. De zwaarste sterren die spiraalstelsels hun blauwe kleur geven hebben slechts een korte levensduur en deze sterren worden in elliptische sterrenstelsels dan ook bijna niet meer aangetroffen. Het ontbreken van deze sterren zorgt ervoor dat elliptische sterrenstelsels over het algemeen geelrood van kleur zijn.

De evolutie van een ster wordt met name bepaald door de massa van een ster. Dit maakt dat de massa van een ster een belangrijke fysische parameter is. Wanneer ergens in het heelal nieuwe sterren gevormd worden, dan gebeurt dit volgens een bepaalde massaverdeling. Deze massaverdeling wordt de initiële massafunctie (IMF) genoemd. In algemene zin kan gesteld worden dat er veel meer lage-massa sterren ($M \leq 0.5$ $M_\odot$[1]) dan hoge-massa sterren ($M > 0.5$ $M_\odot$) worden gevormd. Om te bepalen wat de massaverdeling is van sterren in onze eigen Melkweg, kunnen we in een bepaald volume rondom ons zonnestelsel alle sterren tellen en hun massa

[1]Het symbool $M_\odot$ staat voor zonsmassa.

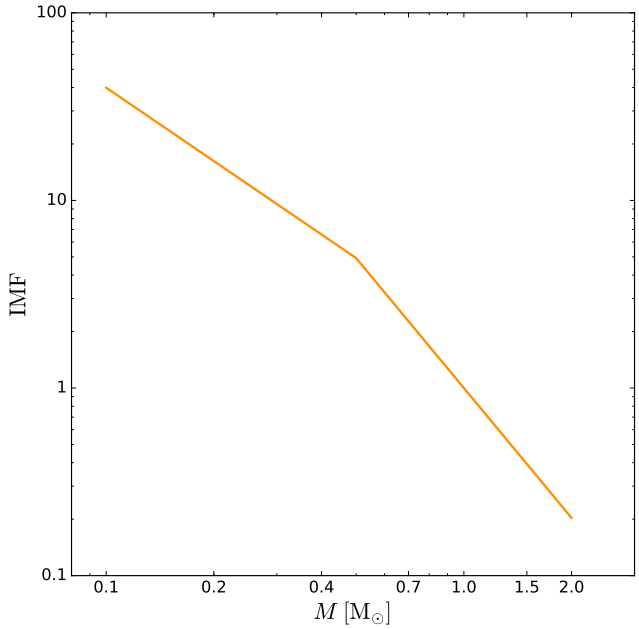

Figure A.2 – Voorbeeld van een massaverdeling van sterren in de Melkweg beschreven aan de hand van twee machtswetten.

bepalen. Dit leidt dan direct tot een stellaire massaverdeling. Hierdoor hebben we een redelijk goed idee van de stellaire massaverdeling in de Melkweg (Salpeter 1955; Kroupa 2001; Chabrier 2003). Het blijkt dat deze massaverdeling voor de Melkweg beschreven kan worden aan de hand van twee machtswetten (Kroupa 2001), zoals te zien in Figuur A.2. Volgens deze massaverdeling, zijn er voor elke ster met een massa van 1 zonsmassa, ongeveer 5 sterren met een massa van 0.5 $M_\odot$ en ongeveer 40 sterren met een massa van 0.1 $M_\odot$.

Voor verderaf gelegen sterrenstelsels is het echter niet mogelijk om de IMF te bepalen aan de hand van stertellingen omdat deze sterrenstelsels slechts een lichtvlekje vormen op onze telescopen en we niet in staat zijn om individuele sterren waar te nemen. Vaak wordt voor de IMF in deze sterrenstelsels aangenomen dat deze hetzelfde is als in de Melkweg. De laatste jaren is er echter steeds meer bewijs gekomen dat dit voor zware elliptische sterrenstelsels niet opgaat en dat zware elliptische sterrenstelsels

meer lage-massa sterren bevatten dan de Melkweg. Deze mogelijke variatie van de IMF kan ons meer vertellen over het proces van stervorming in verschillende sterrenstelsels, maar ook over de vorming en evolutie van deze sterrenstelsels zelf. Daarnaast is er een veelvoud aan sterrenkundige onderzoeken waarbij een aanname wordt gedaan voor de IMF en waarbij deze aangenomen IMF een belangrijke rol speelt. Voorbeelden van dit soort onderzoeken zijn studies waarbij de verhouding tussen massa en licht wordt berekend en gebruikt, studies waarbij de chemische verrijking van sterrenstelsels wordt bestudeerd en studies die betrekking hebben op stellaire overblijfselen zoals supernova's en zwarte gaten. Wanneer de IMF niet universeel is, heeft dit gevolgen voor al deze studies. Het nauwkeurig bepalen van de IMF en eventuele variaties hiervan is dan ook van groot belang voor de sterrenkunde.

In dit proefschrift richt ik mij op het ontwikkelen van een nauwkeurig statistisch model voor het bepalen van de massaverdeling van onopgeloste elliptische sterrenstelsels en op de vraag of deze massaverdeling in alle elliptische sterrenstelsels gelijk is of juist aan verandering onderhevig is. Om de massaverdeling in deze sterrenstelsels te bepalen, wordt gebruik gemaakt van stellaire populatie synthese.

Spectroscopie en stellaire populatie synthese

We hebben reeds gezien dat de kleuren van sterrenstelsels ons iets vertellen over het type sterren dat in een sterrenstelsel voorkomt. Een andere nauwkeurige manier om de verschillende sterren in deze stelsels te bestuderen, is aan de hand van het spectrum van een dergelijk sterrenstelsel. Met behulp van een spectrograaf kan het spectrum van een sterrenstelsel worden bepaald. Hierbij wordt het licht van het sterrenstelsel ontleed, en wordt binnen een bepaald golflengtebereik gekeken wat de lichtintensiteit is als een functie van golflengte. Figuur A.3 toont een voorbeeld van een spectrum van een elliptisch sterrenstelsel in het golflengtegebied van 400 nm tot 900 nm. Zoals te zien, is dit een behoorlijk gecompliceerd spectrum waarin allerlei lijntjes zichtbaar zijn. Een dergelijk spectrum bestaat uit het licht van miljoenen of miljarden verschillende sterren. En ergens in dat spectrum zit de massaverdeling versleuteld waarmee de sterren van dat sterrenstelsel gevormd zijn.

Om te begrijpen hoe het spectrum van een sterrenstelsel tot stand komt, beschouwen we een simpele stellaire populatie (SSP). Dit is een

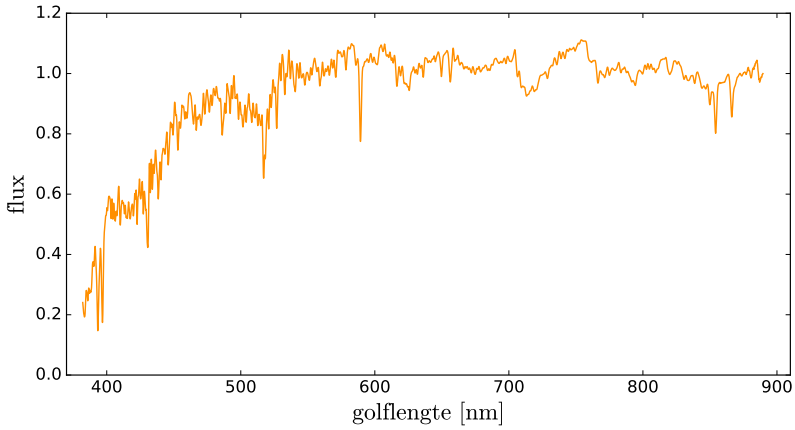

Figure A.3 – Voorbeeld van een spectrum van een elliptisch sterrenstelsel. In het spectrum zien we de lichtintensiteit (flux) als functie van de golflengte.

populatie van sterren die gevormd wordt op hetzelfde moment in tijd en met dezelfde chemische samenstelling. Tijdens de evolutie van een SSP, sterven de zwaarste sterren langzaam uit en wordt de massa van de zwaarste ster die nog in leven is steeds lager. De eigenschappen van een SSP worden dan ook voor een belangrijk deel bepaald door de leeftijd van de SSP, maar ook de chemische samenstelling speelt een belangrijke rol. Aan de hand van numerieke modellen van stellaire evolutie, kunnen sterrenkundigen voorspellen welke sterren er voorkomen in een SSP met een bepaalde leeftijd en chemische samenstelling en wat de eigenschappen zijn van deze sterren (d.w.z. de temperatuur, oppervlaktezwaartekracht en helderheid). In Figuur A.4 is de effectieve temperatuur uitgezet tegen de oppervlaktezwaartekracht en wordt een voorbeeld gegeven van een zogenaamde isochroon. Een isochroon beschrijft een populatie sterren met dezelfde leeftijd. De verschillende punten van de isochroon representeren sterren met verschillende massa's. Een ster brengt het merendeel van zijn leven door op de hoofdreeks. Gedurende deze tijd verbrandt de ster waterstof in zijn kern. Nadat de waterstof in de kern is opgebrand, evolueert de ster van de hoofdreeks naar de reuzenreeks en gaat de ster door verschillende fases van sterevolutie alvorens het uiteindelijk eindigt als witte dwerg, neutronenster of zwart gat (afhankelijk van de massa van de ster). De tijd die een ster doorbrengt op de reuzenreeks is relatief kort

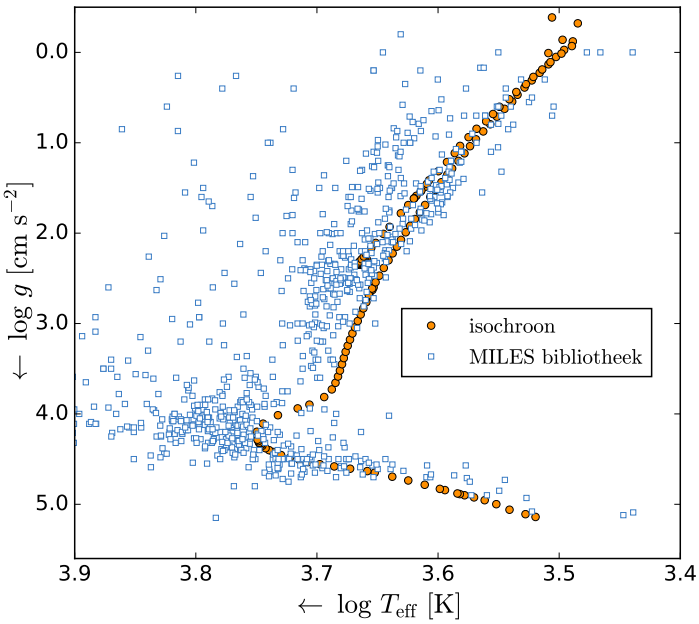

Figure A.4 – De *cirkelvormige oranje* punten vormen samen een isochroon. Deze isochroon is gebaseerd op numerieke modellen van stellaire evolutie en beschrijft alle sterren die voorkomen in een SSP met een leeftijd van 13 miljard jaar en een chemische samenstelling gelijk aan die van de zon. De *blauwe vierkante* punten zijn alle sterren die te vinden zijn in de MILES sterbibliotheek. Door gebruik te maken van een sterbibliotheek kunnen we voor elk van de sterren van de isochroon een spectrum bepalen.

ten opzichte van de tijd doorgebracht op de hoofdreeks. Aan de hand van numerieke modellen van sterevolutie kan bepaald worden hoe een isochroon voor een populatie sterren met een bepaalde leeftijd eruitziet. De isochroon in Figuur A.4 komt overeen met een SSP met een leeftijd van 13 miljard jaar en een chemische samenstelling gelijk aan die van de zon. Het merendeel van de sterren bevindt zich op de hoofdreeks, de zwaarste sterren van de populatie zijn van de hoofdreeks af geëvolueerd en bevinden zich op de reuzenreeksen.

Een isochroon vertelt ons dus welke sterren in een SSP voorkomen maar bevat geen informatie over hoe de spectra van deze sterren eruit zien. Om dit te bepalen, moeten we gebruik maken van een sterbibliotheek. Een sterbibliotheek bevat een grote collectie spectra van sterren met

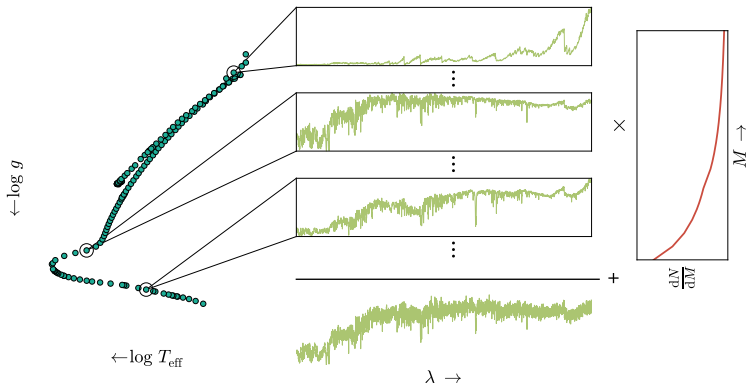

Figure A.5 – Schematische weergave van stellaire populatie synthese van een SSP. Aan de hand van een isochroon, een stellaire bibliotheek en een initiële massafunctie wordt het spectrum van een SSP berekend. Dit spectrum kan vervolgens vergeleken worden met de data om meer te weten te komen over het sterrenstelsel waarvoor we een spectrum tot onze beschikking hebben.

verschillende temperaturen en verschillende stadia van sterevolutie. In Figuur 4 zijn eveneens de parameters van de sterren in de sterbibliotheek MILES (Sánchez-Blázquez et al. 2006) weergegeven. Voor elk van de sterren bevat de bibliotheek een waargenomen spectrum. Door deze waargenomen spectra te interpoleren op de punten van de isochroon, kunnen we in principe voor elk van de sterren in een SSP een spectrum bepalen. Tenslotte kunnen we berekenen hoe het spectrum van de SSP eruit ziet door de spectra van alle sterren in de isochroon bij elkaar op te tellen. Hiervoor ontbreekt echter nog wel één stukje informatie: namelijk hoe vaak elk van de sterren voorkomt in de SSP. Deze laatste stap wordt beschreven door de massaverdeling. De massaverdeling is een waarschijnlijkheidsverdeling die beschrijft hoe vaak elk van de sterren in de SSP voorkomt en deze stelt ons in staat om het spectrum van de SSP uit te rekenen. Dit proces is schematisch weergegeven in Figuur A.5. In werkelijkheid is een sterrenstelsel geen SSP maar kent het een complexe historie van stervorming en bij het modelleren van het spectrum van een sterrenstelsel zal hier rekening mee gehouden moeten worden.

Door isochronen voor verschillende leeftijden en verschillende chemische samenstellingen te combineren met een goede sterbibliotheek en dit te combineren met verschillende aannames voor de IMF, zijn sterrenkundigen

in staat om het spectrum van een sterrenstelsel als het ware te simuleren. Elk van deze gesimuleerde spectra kan vervolgens vergeleken worden met het waargenomen spectrum van het sterrenstelsel en hierdoor zijn we in staat om bijvoorbeeld te bepalen wat de leeftijd is waarop het sterrenstelsel het merendeel van zijn sterren heeft gevormd, maar ook wat de massaverdeling is van de sterren in het sterrenstelsel. Deze techniek waarbij een waargenomen spectrum van een sterrenstelsel wordt getransformeerd in fysische parameters die niet direct waarneembaar zijn, wordt stellaire populatie synthese genoemd. In deze thesis wordt gebruik gemaakt van stellaire populatie synthese om de massaverdeling van sterren in elliptische sterrenstelsels te bepalen.

Dit proefschrift

In dit proefschrift hebben we een nauwkeurig statistisch model ontwikkeld voor het bepalen van de massaverdeling van onopgeloste sterrenstelsels. In klassieke stellaire populatie synthese modellen wordt a priori voor een groot aantal verschillende parameters (zoals SSP leeftijd, chemische samenstelling en IMF) een spectrum berekend en al deze spectra worden vergeleken met het waargenomen spectrum. Op basis hiervan worden dan de beste parameters geselecteerd om het waargenomen sterrenstelsel te beschrijven. In dit proefschrift kiezen we voor een andere benadering waarbij we direct proberen af te leiden welke sterren er in een sterrenstelsel voorkomen. Op basis hiervan bepalen we wat de bijbehorende massaverdeling is.

Een belangrijke eigenschap van het model dat we in dit proefschrift hebben ontwikkeld, is dat het een hiërarchisch Bayesiaans model is. Dat wil zeggen dat het gaat om een statistisch model met verschillende niveaus waarbij de parameters van het model worden vastgesteld op basis van de Bayesiaanse methode. Voor de wiskundige beschrijving van het model verwijs ik naar Hoofdstuk 2 van dit proefschrift. Deze manier van modelleren stelt ons in staat om verschillende ingrediënten van stellaire populatie synthese modellen op een onafhankelijke manier met elkaar te vergelijken. Op basis van het Bayesiaans bewijs kunnen we vervolgens de beste ingrediënten voor ons model selecteren. Daarnaast implementeert deze manier van modelleren automatisch het scheermes van Occam, wat zoveel wil zeggen als dat we bij keuze uit meerdere modellen het simpelste model selecteren dat een goede fit geeft aan de data. Complexere modellen zullen de data altijd beter modelleren maar wanneer de extra parameters

van deze complexere modellen de fit slechts een heel klein beetje verbeteren, is het beter voor een simpeler model kiezen met minder vrije parameters.

Van elliptische sterrenstelsels wordt gedacht dat ze een relatief simpele stervormingshistorie hebben. Daarom wordt vaak aangenomen dat elliptische sterrenstelsels goed gemodelleerd kunnen worden als een simpele stellaire populatie (SSP). In dit proefschrift onderzoeken we de gevolgen van deze aanname en laten we zien dat elliptische sterrenstelsels het best met twee of drie SSP's gemodelleerd kunnen worden. Wanneer slechts één SSP wordt gebruikt, is de helling van de IMF die wordt bepaald te hoog en daarmee wordt de hoeveelheid lage-massa sterren die wordt bepaald overschat. Dit effect verdwijnt wanneer het elliptische sterrenstelsel wordt gemodelleerd met twee of drie SSP's.

In dit proefschrift hebben we de nieuwe stellaire bibliotheek XSL (X-shooter Spectral Library, Chen et al. 2011) gecombineerd met de stellaire bibliotheek MILES tot een nieuwe verzameling van stellaire spectra: de MIX bibliotheek. Deze MIX bibliotheek hebben we gecombineerd met diverse verschillende modelingrediënten (waaronder een variabele chemische samenstelling om te modelleren dat elliptische sterrenstelsels waarschijnlijk een andere chemische samenstelling hebben dan de Melkweg). De combinatie van het statistische model, de MIX library en de verschillende modelingrediënten hebben we toegepast op een verzameling samengestelde spectra van elliptische sterrenstelsels. In totaal hebben we gekeken naar zes samengestelde spectra met verschillende snelheidsdispersies in de reeks van 170 tot 320 km s^{-1}. Aangezien de snelheidsdispersie van een sterrenstelsel gerelateerd is aan de massa van het sterrenstelsels (hogere snelheidsdispersie komt overeen met een zwaarder sterrenstelsels), beschouwen we hiermee sterrenstelsels met verschillende massa's. De resultaten van dit proefschrift komen overeen met eerdere resultaten in de literatuur en laten zien dat de helling van de IMF verandert als functie van de massa. Een belangrijke conclusie van dit proefschrift is dat de IMF van elliptische sterrenstelsels niet gemodelleerd zou moeten worden met een enkele machtswet maar dat gebruik gemaakt zou moeten worden van een dubbele machtswet met twee componenten. Onze resultaten laten zien dat wanneer gebruik wordt gemaakt van een dubbele machtswet om de IMF te modelleren, de hoeveelheid lage-massa sterren ($M \leq 0.5$ $M_\odot$) aan verandering onderhevig is (de zwaarste sterrenstelsels bevatten meer lage-massa sterren). De hoeveelheid hoge-massa sterren ($M > 0.5$ $M_\odot$) is daarentegen ongeveer constant als functie van massa.

Referenties

Chabrier, G. 2003, PASP, 115, 763
Chen, Y., Trager, S., Peletier, R., & Lançon, A. 2011, Journal of Physics Conference Series, 328, 012023
Hubble, E. P. 1936, Realm of the Nebulae
Kroupa, P. 2001, MNRAS, 322, 231
Salpeter, E. E. 1955, ApJ, 121, 161
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703

Acknowledgements

As for many theses, writing this thesis has led to some frustrations now and then (in particular towards the end). But most of the time, doing a PhD and writing this thesis has been quite an enjoyable experience for me. The work that was presented in this thesis would not have been possible without the help, inspiration and support of many people and in the last pages of this thesis I would like to express my gratitude to those people.

First and foremost, I would like to thank my promotors. Scott and Leon, I have always felt very honored to be your student and both of you have been a great supervisor to me. I think that the two of you form an excellent team, and the knowledge and experience that the two of you have, has been perfectly complementary for my project and my development as a scientist. Scott, your inexhaustible knowledge of astronomy and your contagious enthusiasm for this branch of science make you a very good teacher. During these four years (but also before), you have taught me a lot about astronomy and being a scientist. Although I will not continue to work in science for now, I will definitively continue to think like a scientist and in that respect some of the things that I learned over the last four years will be useful for the rest of my life. As a person, you have always been very friendly, patient and understanding, which makes working with you a real pleasure. Leon, during my bachelor and master I always felt that I kind of understood what Bayesian statistics was about. You showed me that I was wrong and in the past four years you have taught me a lot about Bayesian model comparison. Now I am again in a position where I think that I kind of understand Bayesian statistics, but based on previous experience I admit

that there is a nonzero probability that I do not completely understand it. You have always been very critical to my scientific writing and in that way you have taught me a lot about writing a paper and presenting my ideas to the scientific community. Your friendliness and sense of humour make working with you a very nice experience.

Special thanks go out to the members of the XSL team. Alexandre, Anaïs, Anke, Ariane, Jesus, Kristiina, Mariya, Mathieu, Mina, Omar, Patricia, Philippe, Reynier and Yanping, we have shared many telecons and have met in person during one of the busy weeks in Groningen, Strassbourg or Tenerife. I really enjoyed being part of this team and working together with all of you. Anaïs and Mariya, you allowed me to be a small link in the data reduction process of XSL where I worked on the telluric corrections of the spectra. Thank you for what has been a very enjoyable collaboration for me. Moreover, I would like to thank the two of you for the interesting scientific and non-scientific conversations that we shared together. Reynier, thank you for always being interested in my work, for asking many questions and also for being critical to my work.

Gergö and Rachel, thank you for working together with me on my second scientific paper. Chiara, thanks a lot for the collaboration that led to Chapter 4 of this thesis, for always being curious about my work and for sharing your enthusiasm about astronomy and in particular the IMF. Anaïs, Anke, Gergö, Gerjon, Judith, Kristiina, Mariya, Shoko and Wouter, I would like to thank you for all the scientific discussions, for sharing your work with me and for commenting on my work during the (more or less regular) weekly group meetings.

I am very grateful to the reading committee, Charlie Conroy, Reynier Peletier and Tomasso Treu, for accepting the task to read my thesis.

Although office 140 does not exist anymore, I consider myself very lucky to have been sharing this nice office for more than three years with Marisa, Smaran and Wouter. Wouter, I really enjoyed the conversations that we had and the fact that you are critical about almost everything. I would also like to thank you for your comments on the Dutch summary. Marisa, you have been there during almost the entire course of my PhD and I am very happy that you have been my office mate for all those years. Thank you for all the interesting conversations that we shared, for sharing your experiences about life in Italy, for always being very friendly to me and showing your interest in my work and life. Smaran, we did not share an

office for a long time but I have come to know you as a friendly person and would have been happy to have shared the office for a longer period of time.

I would like to thank the Computer Group, Eite, Leon, Martin, Valentin and Wim, for answering my questions and for always quickly solving any system-related problems that I encountered. A university can now and then be quite bureaucratic and I would like to thank the secretaries Christa, Gineke, Gonda, Hennie, Martine and Ramona for making sure that I mostly had to worry about my PhD and not about anything else. Lucia, thanks for showing interest in the progress of my thesis and the search for a new job in my final year.

The Kapteyn Institute has always provided a very pleasant working environment and during my PhD time many different people contributed to the fact that the Kapteyn Institute is such a nice place. I am deliberately not going to mention any specific names here, since I will most likely forget to mention someone. But if you have been at the Kapteyn Institute for the past four years, you have most likely contributed to making it a nice place and I would like to thank you for that.

I would like to thank my friends Bianca, Cathrin, Erwin, Hessel, Jantine, Karlijn, Lammert, Marcel, Marcel, Merel, René, Renske, Rianne, Rik and Ruud for providing the necessary distraction from my thesis over the last four years. With some of you, I have spent most of my Wednesday nights over the last years and sharing a pizza together and playing a game has always been a very enjoyable and relaxing moment for me.

Aske, Renze, Erwin, Frouke, Gijsbert en Hayo, bedankt voor de openheid en warmte waarmee ik binnen jullie familie ben ontvangen en voor jullie interesse in mijn onderzoek en mijn leven. Graag wil ik mijn familie, Remke, Ida, Alex en Hester bedanken voor de getoonde interesse, ondersteuning en mooie momenten die we samen gedeeld hebben. Susan en Teun, ik waardeer het enorm dat jullie speciaal voor mijn verdediging zijn teruggekomen van jullie grote reis. Bedankt voor alle leuke dingen die we de afgelopen jaren samen gedaan hebben. Susan, dank voor het controleren van mijn Nederlandse samenvatting. Pap en mam, ik ben jullie enorm dankbaar dat jullie ervoor gezorgd hebben dat ik zorgeloos heb kunnen opgroeien, dat jullie altijd vertrouwen hebben gehad in de keuzes die ik gemaakt heb, voor jullie interesse en voor de geboden ondersteuning. Zonder jullie was deze thesis nooit tot stand gekomen. Lieve, lieve Caroline, wat was het toch fijn dat ik de afgelopen jaren met jou heb mogen delen. Jij bent het beste dat me ooit overkomen is!